

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/12, C07K 14/47		A1	(11) International Publication Number: WO 99/63080
			(43) International Publication Date: 9 December 1999 (09.12.99)
(21) International Application Number: PCT/EP99/03848		CRAEN, Marc [-/BE]; Devgen N.V., Blok DF 1.60.14, Technologiepark Zwijnaarde 9, B-9052 Zwijnaarde (BE).	
(22) International Filing Date: 2 June 1999 (02.06.99)		(74) Agents: MAYES, Stuart, David et al.; Boulton Wade Tennant, 27 Furnival Street, London EC4A 1PQ (GB).	
(30) Priority Data: 9811962.1 3 June 1998 (03.06.98) GB		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(71) Applicant (for all designated States except US): JANSSEN PHARMACEUTICA N.V. [BE/BE]; Turnhoutseweg 30, B-2340 Beerse (BE).		Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(72) Inventors; and (75) Inventors/Applicants (for US only): LUYTEN, Walter, Herman, Maria, Louis [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). DE RAEYMAEKER, Marc, Carl [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). GEYSEN, Johan, Jozef, Gustave, Hendrik [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). BOGAERT, Thierry, A., O., E. [BE/BE]; Wolvendreef 26G, B-8500 Kortrijk (BE). MAERTEN, Luc, Jacques, Simon [BE/BE]; Devgen N.V., Blok DF 1.60.14, Technologiepark Zwijnaarde 9, B-9052 Zwijnaarde (BE). VERHASSELT, Peter [-/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). VAN DE			
(54) Title: HUMAN HOMOLOGUE OF UNC-53 PROTEIN OF <i>C. ELEGANS</i>			
(57) Abstract <p>There is disclosed human homologues of the UNC-53 protein of <i>C. elegans</i> and cDNA sequences coding for said homologues or functional equivalents thereof. The invention also relates to processes for identifying compounds which control cell behaviour, compounds identified and pharmaceutical compositions containing them in addition to processes and assays for identifying disease states in which said gene or protein is dysfunctional.</p>			

BEST AVAILABLE COPY

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

- 1 -

HUMAN HOMOLOGUE OF UNC-53 PROTEIN OF *C. ELEGANS*

5 The present invention relates to a vertebrate
homologue of UNC-53 protein of *C. elegans* and cDNA
sequences coding for said homologues or functional
equivalents thereof. The invention also relates to
processes for identifying compounds which control cell
behaviour, compounds identified and pharmaceutical
10 compositions containing them in addition to processes
and assays for identifying disease states in which
said gene or protein is dysfunctional.

The control of cell motility, cell shape and
directionality of cell outgrowth of axones or other
15 cell outgrowths is an essential feature in the
morphogenesis and function of both unicellular and
multicellular organisms.

Some cell surface proteins and extra-cellular
molecules controlling the directionality and potential
20 of cell migration have been identified, although the
processes involved are not generally understood. It
is generally considered that a long-range migration of
a cell process (also known as a growth cone extension)
is a stepwise event, whereby prior to and after each
25 extension there is the formation of a structure at the
leading edge of the cell. Localised stabilisation of
the actin cytoskeleton and association with plus end
regions of microtubules is a general cell biological
process underlying the choice of directional
30 extension.

The present inventors have surprisingly found a
new human gene/protein belonging to the UNC-53 family
that binds microtubules and, in particular, the plus-
end regions of microtubules.

35 A gene from the free-living nematode
Caenorhabditis elegans designated "unc-53" has been
previously identified and cloned (Abstract,

- 2 -

International C. elegans Meeting, June 1-5 1991,
Madison, Wisconsin, 58, Bogaert and Goh). The present
inventors previously identified UNC-53 protein as a
signal transducer or signal integrator controlling the
5 directionality of cell migration and/or cell shape in
C. elegans (WO 96/38555).

The C. elegans UNC-53 protein (Ceunc53) and
previously found human homologues thereof (hs-unc53/1
and hs-unc53/2) were found to encode a signal
10 transducer or a signal integrator, controlling the
directionality of a cell migration, cell shape and
growth extension. Evidence indicates that the
presently found homologue designated (hs-unc53/3)
might act as an adapter linking extracellular signals
15 to the actin cytoskeleton. Firstly hs-unc-53/3 shows
homology to the cortical actin binding proteins, and
the Ce-UNC-53 protein has been shown to bind F-actin
in vitro and leads to actin re-organization in vivo
when expressed in mammalian cells, leading to an
20 increased number of filopodia and lamellipodia.
Furthermore, increased neurite extension and increased
cell motility could be observed. Hs-UNC-53-3 may play
an important role in the development of various
diseases.

25 According to a first aspect of the present
invention there is provided a vertebrate protein
homologue of an UNC-53 protein of C. elegans, which
protein comprises an amino acid sequence having one or
more of sequence blocks A, B, C, D, E, F, G or H as
30 illustrated in figure 4 or which differs from said
blocks in conservative amino acid changes.

According to a further aspect of the present
invention, there is provided a vertebrate protein
homologue of UNC-53 protein of C. elegans or a
35 functional equivalent, derivative or bioprecursor
thereof, having an amino acid sequence encoded by the
nucleotide sequence illustrated in figure 1(e).

- 3 -

For the purposes of the present invention a "derivative" should be taken to mean mutational derivatives, fusions, internal deletions, splice variants and muteins.

5 Preferably, said vertebrate homologue is a human protein, and preferably a mammalian or a mouse protein.

A further aspect of the invention comprises a vertebrate homologue comprising an amino acid sequence as shown in figure 1(f) or the variants thereof or an amino acid sequence which differs from the amino acid sequences shown in figure 1(f) to a significant extent only in one or more conservative amino acid changes.

10 In a further aspect of the present invention there is also provided a nucleic acid molecule, which is preferably DNA, and which encodes a vertebrate homologue of UNC-53 protein of C. elegans, or a functional equivalent derivative, fragment or bioprecursor of said homologue according to the invention. Preferably, the cDNA comprises a sequence of nucleotides encoding an amino acid sequence as illustrated in figure 1(f) or the variants thereof or an amino acid which differs from the sequences shown in these figures to a significant extent only in one or more conservative amino acid changes. Preferably the DNA is cDNA, which cDNA comprises the sequence shown in figure 1(e) or the variants indicated therein. Also provided by the present invention is a nucleic acid sequence capable of hybridising to the nucleic acid or DNA sequences according to the invention under high stringency conditions, which conditions are well known to those skilled in the art.

15 The cDNA according to the invention may be included in an expression vector which may itself be used to transform or transfect a host cell, which cell may be bacterial or eukaryotic in origin including such as, for example an animal or plant cell a fungal

- 4 -

cell or an insect cell. Thus, advantageously, once the cDNA corresponding to the genome of the vertebrate homologue of UNC-53 of C. elegans according to the invention is synthesised, using for example, reverse transcriptase or the like, a range of cells, tissues or organisms may be transfected following incorporation of the selected cDNA clone into an appropriate expression vector. The expression vector according to the invention may comprise a promoter of C. elegans or one of human, mouse or viral origin and optionally a sequence encoding a reporter molecule, such as, for example, green fluorescent protein.

The present invention, therefore, also further comprises a transgenic cell, tissue or organism comprising a transgene capable of expressing a vertebrate homologue of UNC-53 protein of C. elegans according to the invention. The term "transgene capable of expressing a vertebrate homologue of UNC-53 protein of C. elegans" as used herein means a suitable nucleic acid sequence which leads to the expression of a vertebrate homologue of UNC-53 protein of C. elegans according to the invention having the same function and/or activity. The transgene may include, for example, genomic nucleic acid isolated from the appropriate vertebrate or synthetic nucleic acid including cDNA. The term "transgenic organisms, tissues or cells, as used herein means any suitable organism and/or part of an organism, tissue or cell, that contains exogenous nucleic acid either stably integrated in the genome or in an extrachromosomal state.

Preferably the transgenic cell comprises any of, a COS cell, HepG2 cell, MCF-7 or N4 neuroblastoma cell, a NIH3T3 cell, a colorectal or carcinoma cell or a human derived cell such as a fibroblast or the like. The transgenic organism may be an insect, a non-human animal or a plant and preferably C. elegans or a

- 5 -

related nematode. Preferably, the transgene comprises the nucleic acid or cDNA sequence encoding the vertebrate homologue according to the invention as described above. The transgene preferably comprises an expression vector according to the invention.

The term "functional fragment" as used herein should be taken to mean a fragment of the gene coding for the vertebrate homologue of the UNC-53 protein of C. elegans according to the invention. For example, the gene may comprise deletions or mutations but may still encode a functional vertebrate homologue of UNC-53 protein.

Further provided by the present invention is a method of producing a mutant vertebrate non-human organism having a mutation in the wild-type gene coding for the vertebrate homologue of UNC-53 protein according to the invention, which mutation affects cell behaviour or the regulation of cell motility or the shape or the direction of cell migration or microtubule plus end stability or function and localisation of protein complexes located thereon, which method comprises inducing a mutation in the vertebrate homologue of UNC-53 protein in said organism. These mutant organisms may be used in a screen to identify the effects of compounds on these cell functions.

The vertebrate homologue of UNC-53 protein of C. elegans or the cDNA or genomic DNA encoding it or a functional equivalent, derivative, fragment or bioprecursor of said homologue, may advantageously be used as a medicament, or in the preparation of a medicament to treat or prevent disorders associated with inhibition of overexpression of the vertebrate homologue of UNC -53 according to the invention. Such disorders may be alleviated by promoting neuronal regeneration, revascularisation or wound healing or the treatment of chronic neurodegenerative disorders,

- 6 -

psychiatric disorders or acute traumatic injuries or fibrotic disease or disease in which physiological events requiring the polarity of cells or epithelia are abnormally functioning. Accordingly, the

5 vertebrate homologue according to the invention, dominant positive or negative mutants thereof, or inhibitors thereof may advantageously be used to induce or alleviate contact inhibition in a cell or in preventing carcinoma development. Typically, the

10 above medical conditions may be treated in mammals and more preferably humans by either the homologue of UNC-53 protein or alternatively by a nucleic acid coding for the protein or the protein itself according to the invention. Alternatively an antisense oligonucleotide

15 to said UNC-53 vertebrate homologue may be used to prevent its expression. Examples of other nucleic acid sequences which may be used include 3' untranslated regions of mRNA which could be used to prevent transcription of the genomic sequence encoding

20 for the vertebrate homologue of UNC-53 protein according to the invention.

The vertebrate homologue of UNC-53 protein according to the invention may be incorporated into a pharmaceutically acceptable composition together with

25 a suitable carrier, diluent or excipient therefor. The pharmaceutical composition may advantageously comprise, additionally or alternatively, the nucleic acid sequence according to the invention as defined above.

30 The induction or inhibition of the expression of hu-UNC-53/3 by pharmacological means may advantageously be used to induce neuronal regeneration, revascularisation or wound healing or be involved in the treatment of chronic

35 neurodegenerative disorders, or acute traumatic injuries or fibrotic diseases, or physiological events requiring the polarity of cells, or oncology and

- 7 -

metastasis of cells, or apoptotic pathways.

The present invention therefore also provides for a method of determining whether a compound is an inhibitor or enhancer of the regulation of cell
5 behaviour, growth, transformation, cell shape or motility or the direction of cell migration, microtubule plus end stability or function and localisation of protein complexes thereon, which method comprises contacting said compound with a
10 transgenic cell according to the invention and screening for a phenotypic change in said cell. The method can therefore be used to determine whether the compound comprises an inhibitor or an enhancer of the signal transduction pathway of said transgenic cell of
15 which pathway said vertebrate homologue of UNC-53 protein according to the invention is a component, or whether said compound is an inhibitor or an enhancer of a parallel or redundant signal transduction pathway in said cell. The present invention also provides a
20 method to determine that the protein in said signal transduction pathway is a vertebrate homologue of UNC-53 protein of C. elegans according to the invention.

Preferably, the phenotypic change to be screened comprises a change in cell shape or a change in cell
25 motility. Where a transgenic cell is used in accordance with one embodiment of the method of the invention, an N4 neuroblastoma cell may be used and in such an embodiment the phenotypic change to be screened may be the length of neurite growth, changes
30 in filopodia outgrowth, changes in ruffling behaviour or cell adhesion, any change in microtubule cytoskeleton, any change in localisation of proteins on plus end regions of microtubules or any change in a cell such as apoptosis. In an alternative embodiment
35 of the method of the invention, the transgenic cell may comprise an MCF-7 breast carcinoma cell. Typically in such an embodiment the phenotypic change

- 8 -

to be screened comprises the extent of phagokinesis or
filopodia formation. In an alternative embodiment of
this aspect of the invention, the transgenic cell may
comprise an NIH3T3 cell. Typically in such an
5 embodiment the phenotypic change to be screened
comprises loss of contact inhibition of foci
formation. The method according to the invention, may
also utilise a mutant cell or mutant organism
according to the invention as described above, where
10 the mutant cell is capable of growing in tissue
culture or in vivo and either of which cell or
organism has a mutation in the wild-type unc-53 gene.

In accordance with the present invention, a
"phenotypic change", may comprise any phenotype
15 resulting from changes at any suitable point in the
life cycle of the cell, tissue or organism defined
above, which change can be attributed to the
expression of the transgene of the invention such as
for example, growth, viability, morphology, behaviour,
20 movement, cell migration or cell process or growth
cone extension of cells and includes changes in body
shape, locomotion, chemotaxis, contact inhibition,
mating behaviour or the like. The phenotypic change
may preferably be monitored directly by visual
25 inspection of the cell as a whole or by monitoring the
F-actin cytoskeleton microtubule network and plus end
stability of microtubules or proteins thereon or
alternatively by for example measuring indicators of
viability including endogenous or transgenically
30 introduced histochemical markers or other reporter
genes, such as for example β -galactosidase or green
fluorescent protein.

A compound which is identifiable by the method
according to the invention as described above, as an
35 enhancer of the processes identified above such as the
regulation of cell shape or motility or the direction
of cell migration may be used as a medicament, or

- 9 -

alternatively in the preparation of a medicament, for promoting neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neuro-degenerative diseases or acute traumatic injuries or fibrotic disease. Examples of promoting neuronal regeneration include, for example, peripheral nerve regeneration after trauma and spinal cord trauma.

Where a compound is identified in accordance with the method described above as being an inhibitor of the regulation of cell shape or mobility or the direction of cell migration, the compound may be used as a medicament, or in the preparation of a medicament, for substantially alleviating spread of disease inducing cells, such as in spread of carcinoma, or the like in metastasis or in alleviating loss of contact inhibition. Advantageously, any of the compounds which may have been identified as an inhibitor or an enhancer in accordance with the method as described above, may also be included in a pharmaceutical composition comprising the respective compound and a pharmaceutically acceptable carrier, diluent or excipient therefor.

The particular mechanism of action of a compound identified as either an inhibitor or an enhancer of the cell motility shape, growth or direction of cell migration or microtubule association or to the plus end region thereof is not limiting. Preferably the compound acts as an inhibitor or enhancer of a signal transduction pathway. The compound may also act on a parallel pathway or directly on the vertebrate homologue of UNC-53 protein of C. elegans. For example, the method of action of the compound may include direct interaction with the vertebrate homologue of UNC-53 protein, interaction with processes for regulating phosphorylation or dephosphorylation of the vertebrate homologue of UNC-53 or with processes regulating activity of an unc-53

- 10 -

gene or with processes for post-transcriptional or post-translational modification or the like.

Preferably the compound is identified by the method according to the invention as an inhibitor or an enhancer, by utilising differences of phenotype of the cell, tissue or organism, which are visible to the eye. Alternatively indicators of viability including endogenous or transgenically introduced histochemical markers or a reporter gene may be used.

According to a further aspect of the invention there is also provided a transgenic cell or tissue culture which has been constructed to comprise a promoter sequence of a gene coding for a vertebrate homologue of UNC-53 of C. elegans according to the invention operably linked to a nucleic acid sequence encoding a reporter molecule. Preferably, the reporter sequence encodes for a detectable protein, for example one which may be monitored by eye inspection such as antibiotic resistance, β -galactosidase or a molecule detectable by spectrophotometric, spectrofluorometric, luminescent or radioactive assays.

The present invention also provides a method of determining whether a compound is an inhibitor or an enhancer of transcription of a gene coding for a vertebrate homologue of UNC-53 protein in C. elegans, according to the invention which method comprises the steps of:

- (a) contacting said compound with a transgenic cell according to the invention as described above,
- (b) monitoring the level of said reporter molecule and comparing results obtained from this monitoring step with a control comprising a transgenic cell having the promoter sequence of a gene coding for a vertebrate homologue of UNC-53 protein, or a functional fragment of said

- 11 -

homologue and the reporter molecule, in the absence of the compound.

In one embodiment of the method according to this aspect of the invention the reporter molecule may
5 comprise messenger RNA.

A compound identified as an enhancer of transcription of the gene coding for the vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, derivative or bioprecursor of
10 said homologue may also be used as a medicament, or in the preparation of a medicament, for promoting neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neuro-
degenerative diseases or acute traumatic injuries or
15 fibrotic disease. Furthermore, such compounds may be included in a pharmaceutical composition including a pharmaceutically acceptable carrier, diluent or excipient therefor. Any compounds identified as
inhibitors of transcription may, advantageously, be
20 used in alleviating the spread of disease inducing cells such as carcinomas or metastasis or loss of contact inhibition.

The present invention also provides a kit for determining whether a compound is an enhancer or an
25 inhibitor of the regulation of cell growth, transformation, cell motility or shape or the direction of cell migration which kit comprises at least one transgenic or mutant cell or transgenic or mutant non-human organism according to the invention
30 as described above and a plurality of wild-type cells or a wild-type organism of the same type, or a cell line or tissue culture and means for contacting said compound with said cell or organism.

Also provided by the present invention is a kit
35 for determining whether a compound is an inhibitor or an enhancer of transcription of a gene coding for a vertebrate homologue of UNC-53 protein of C. elegans

- 12 -

according to the invention which kit comprises at least one transgenic cell or cells according to the invention, means for contacting said compounds with said cells and means for monitoring the level of transcription of said transgenic cell or cells according to the invention.

For the purposes of the present invention, the term "gene coding for a vertebrate homologue of UNC-53 or a functional fragment of said homologue" includes the nucleic acid sequence shown in figure 1 or a fragment thereof, including the differentially spliced isoforms and transcriptional starts of the nucleic acid sequence and which sequence encodes a vertebrate homologue of UNC-53 protein or a functional equivalent, derivative, fragment or bioprecursor of the protein.

The present invention also provides methods of identifying genes of vertebrates or fragments of said genes, which encode proteins which are active in the signal transduction pathway of which the vertebrate homologue of UNC-53 according to the present invention is a component. A preferred method comprises hybridizing to an appropriate cDNA library a nucleotide sequence, as defined herein, or a fragment thereof under appropriate conditions of stringency in order to identify genes having statistically significant homology with the cDNA clones of any one of the cDNA sequences according to the invention described above.

Furthermore, there is also provided by the present invention a method of identifying a protein which is active in the signal transduction pathway of a cell of which a vertebrate homologue of UNC-53 protein of C. elegans according to the invention is a component. According to this aspect of the invention, the method comprises;

(a) contacting an extract of said cell with an

- 13 -

antibody to the vertebrate homologue of UNC-53 protein or a functional equivalent, fragment or bioprecursor of said protein,

- 5 (b) identifying the antibody/vertebrate homologue of UNC-53 complex, and
(c) analysing the complex to identify any protein bound to the vertebrate homologue of UNC-53 protein other than the antibody.

10 The vertebrate homologue of UNC-53 protein, therefore may bind regions of other proteins involved in the signal transduction pathway. It is also possible to sequentially identify a whole range of proteins involved in the signal transduction pathway.

15 Antibodies to the vertebrate homologue of UNC-53 protein may be produced according to known techniques as would be known to those skilled in the art. For example, polyclonal antibodies may be prepared by inoculating a host animal, such as a mouse, with a protein or epitope of a protein according to the
20 invention and recovering immune serum.

This aspect of the invention, further comprises a method of identifying a further protein or proteins which are active in the signal transduction pathway of a cell of which the vertebrate homologue of UNC-53 is
25 a component which method comprises:

- (a) forming an antibody to the first identified protein bound to the vertebrate homologue of UNC-53 protein in the method as described above,
(b) contacting a cell extract with the antibody,
30 (c) identifying any antibody/protein complex,
(d) analysing the complex to identify any further protein bound to the first protein other than the antibody, and
(e) optionally repeating steps (a) to (d) to
35 identify further proteins in the pathway.

According to this aspect of the present invention, the antibody starts the process by binding

- 14 -

to the vertebrate homologue of UNC-53 protein according to the invention in the signal transduction or oncogenic pathways. Any other proteins found complexed to the bound antibody or UNC-53 protein can then be used to identify further interacting proteins involved in the pathway.

It may also be possible to identify proteins involved in the signal transduction pathway of a cell of which the vertebrate homologue of UNC-53 is a component by using a vertebrate homologue of UNC-53 protein of C. elegans. According to this aspect of the invention the method comprises:

- (a) contacting an extract of the cell with the vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, fragment or bioprecursor of said homologue,
- (b) identifying the vertebrate homologue of UNC-53 protein/protein complex formed and
- (c) analysing the complex to identify any protein bound to the vertebrate homologue of UNC-53 protein other than the same vertebrate homologue of UNC-53 protein.

This method can also advantageously be used to identify further proteins in a signal transduction pathway of a cell by contacting an extract of the cell used as described above, with any protein identified from step (c) above not being a vertebrate homologue of UNC-53 protein and repeating steps (b) and (c).

Other methods which may be used for identifying proteins in a signal transduction pathway of a cell may comprise for example a western blot overlay method which method is well known to those skilled in the art. Cell extracts are run on gels to separate out protein and subsequently blotted onto a nylon membrane. These membranes may then be incubated, for example in a medium containing vertebrate homologue of UNC-53 having a label attached thereto such as a

- 15 -

biotin or radiolabel and any protein conjugates visualised with for example a streptavidin or alkaline phosphatase conjugated antibody.

5 The present invention also advantageously provides a process for the preparation of binding antibodies which recognise proteins or fragments thereof involved in the rate and direction of cell migration or the control of cell growth or shape, for the above methods.

10 The monoclonal antibody for binding to the appropriate vertebrate homologue of UNC-53 (or its functional equivalent) may be prepared by known techniques as described by Kohler R. and Milstein C., (1975) Nature 256, 495 to 497.

15 Another method which may be used to identify proteins involved in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein of C. elegans according to the invention or is a component, involves investigating protein-protein
20 interactions using the two-hybrid vector method. This method, which is well known to those skilled in the art was first developed in yeast by Chien et al (1991). This technique is based on functional reconstruction in vivo of a transcription factor which
25 activates a reporter gene. More particularly the technique comprises providing an appropriate host cell with a DNA construct comprising a reporter gene under the control of a promoter regulated by a transcription factor having a DNA binding domain and an activating
30 domain, expressing in the host cell a first hybrid DNA sequence encoding a first fusion of a fragment or all of a nucleic acid sequence according to the invention and either said DNA binding domain or said activating domain of the transcription factor, expressing in the
35 host at least one second hybrid DNA sequence, such as a library or the like, encoding putative binding proteins to be investigated together with the DNA

- 16 -

binding or activating domain of the transcription factor which is not incorporated in the first fusion; detecting any binding of the proteins to be investigated with a protein according to the invention
5 by detecting for the presence of any reporter gene product in the host cell; optionally isolating second hybrid DNA sequences encoding the binding protein.

An example of such a technique utilises the GAL4 protein in yeast. GAL4 is a transcriptional activator
10 of galactose metabolism in yeast and has a separate domain for binding to activators upstream of the galactose metabolising genes as well as a protein binding domain. Nucleotide vectors may be constructed, one of which comprises the nucleotide
15 residues encoding the DNA binding domain of GAL4. These binding domain residues may be fused to a known protein encoding sequence, such as for example a sequence coding for the vertebrate homologue of UNC-53. The other vector comprises the residues
20 encoding the protein binding domain of GAL4. These residues are fused to residues encoding a test protein, preferably from the signal transduction pathway of the vertebrate in question. Any interaction between the vertebrate homologue of UNC-53 protein and
25 the protein to be tested leads to transcriptional activation of a reporter molecule in a GAL-4 transcription deficient yeast cell into which the vectors have been transformed. Preferably, a reporter molecule such as β -galactosidase is activated upon
30 restoration of transcription of the yeast galactose metabolism genes. This method enables any interactions between proteins involved in the signal transduction pathway or a parallel or redundant pathway to be investigated.

35 Any proteins identified in the signal transduction pathway of the cell, which may be for example a mammalian cell, may also be included in a

- 17 -

pharmaceutical composition together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

5 The present invention also provides a process for producing a vertebrate homologue of an UNC-53 protein of C. elegans according to the invention which process comprises culturing the cells transformed or transfected with a cDNA expression vector having any of the cDNA sequences according to the invention as
10 described above, and recovering the expressed protein homologue. The cell may advantageously be a bacterial, animal, insect or plant cell.

A particularly preferred process for producing said vertebrate homologue of UNC-53 protein uses
15 insect cells. Accordingly, the invention provides a process for producing a vertebrate homologue of UNC-53 protein of C. elegans according to the invention which process comprises culturing an insect cell transformed or transfected with a recombinant Baculovirus vector,
20 said vector comprising a nucleotide sequence encoding said vertebrate homologue of UNC-53 protein according to the invention downstream of the Baculovirus polyhedrin promoter and recovering the expressed protein. Advantageously, this method produces large
25 amounts of protein for recovery. The insect cell may be from for example Spodoptera frugiperda or Drosophila Melanogaster.

In accordance with the present invention, a defined nucleic acid sequence includes not only the
30 identical nucleic acid but also any minor base variations from the natural nucleic acid sequence including in particular, substitutions in bases which result in a synonymous codon (a different codon specifying the same amino acid), due to the degenerate
35 code in conservative amino acid substitution. The term "nucleic acid sequence" also includes the complimentary sequence to any single stranded sequence

- 18 -

given which includes the definition above regarding base variations.

Furthermore, a defined protein, polypeptide or amino acid sequence according to the invention, includes not only the identical amino acid sequence but also minor amino acid variations from the natural amino acid sequence including conservative amino acid replacements (a replacement by an amino acid that is related in its side chains). Also included are amino acid sequences which vary from the natural amino acid but result in a polypeptide which is immunologically identical or similar to the polypeptide encoded by the naturally occurring sequence. Such polypeptides may be encoded by a corresponding nucleic acid sequence.

A further aspect of the invention provides a nucleic acid sequence of at least 15 nucleotides of a nucleic acid according to the invention and preferably from 15 to 50 nucleotides.

These sequences may, advantageously be used as probes or primers to initiate replication or the like. Such nucleic acid sequences may be produced according to techniques well known in the art, such as by recombinant or synthetic means. They may also be used in diagnostic kits or the like for detecting for the presence of a nucleic acid according to the invention. These test generally comprise contacting the probe with a sample under hybridising conditions and detecting for the presence of any duplex formation between the probe and any nucleic acid in the sample. Nucleic acid sequences according to the invention may also be produced using recombinant or synthetic means such as described in Sambrook et al (Molecular Cloning: A Laboratory Manual, 1989). Advantageously, human allelic variants or polymorphisms of the DNA according to the invention may be identified by, for example, probing DNA from a range of individuals for example from different populations. Furthermore,

- 19 -

nucleic acids and probes according to the invention may be used to sequence genomic DNA from patients using techniques well known in the art, such as the Sanger Dideoxy chain termination method, which may
5 advantageously ascertain any predisposition of a patient to certain disorders.

A method of detecting whether a compound is an inhibitor or an enhancer or expression of a vertebrate homologue of UNC-53 of C. elegans, according to the
10 invention is also provided which method comprises contacting a cell expressing said homologue with said compound and monitoring for a phenotypic change compared to a control cell which has not been contacted with said compound.

15 Preferably the cell is a transgenic cell as described above. Alternatively the cell may have undergone loss of contact inhibition.

The present method also provides for determining whether said compound is an inhibitor or expression of
20 said vertebrate homologue. In one embodiment the compound to be tested comprises a nucleic acid.

Preferably said nucleic acid sequence comprises an antisense DNA sequence or a mRNA sequence.

25 Preferably said mRNA sequence comprises 3' untranslated regions of mRNA encoding for said vertebrate homologue.

Alternatively, the compound to be tested may be a protein. Preferably, said protein comprises a protein having an amino acid sequence potentially suitable for
30 inhibiting function of said vertebrate homologue and preferably comprises a protein identified by the methods as described herein.

The present invention also provides a pharmaceutical composition comprising a compound, for
35 example an antisense nucleic acid identified according to the above described method together with a pharmaceutically acceptable carrier, diluent or

- 20 -

excipient therefor.

A nucleic acid sequence or protein identified according to this aspect of the invention may be used as a medicament, or in the preparation of a
5 medicament, for treating loss of contact inhibition of cancer which is mediated by vertebrate homologue of UNC-53 protein or a functional equivalent, fragment, derivative or bioprecursor of said homologue.

Further provided by the invention is a nucleic
10 acid as defined above for use in preparation of a medicament for inhibiting expression of a gene coding for a vertebrate homologue of UNC-53 protein of C. elegans.

Further provided by the invention is an assay for
15 detecting expression of the vertebrate homologue of UNC-53 protein of C. elegans in a vertebrate cell which assay comprises contacting a cell or an extract thereof with an antibody to said vertebrate homologue, which antibody is fused to a reporter molecule,
20 removing any unbound antibody and monitoring for the presence of said reporter molecule.

Preferably the reporter molecule is an antibody conjugated to for example a fluorophore such as fluorescein or alternatively to an enzyme such as
25 strepavidin.

There is also provided a method for detecting for expression of a gene coding for the vertebrate homologue of UNC-53 protein of the invention which method comprises contacting a probe specific for a
30 nucleic acid or protein sequence coding for or corresponding to said vertebrate homologue according to the invention with a cell extract, which probe is linked to a reporter and analysing for the presence of said reporter.

35 Preferably the probe is a complementary sequence to a region of mRNA transcribed from said gene encoding said vertebrate homologue of UNC-53 protein

- 21 -

according to the invention.

Preferably the complimentary sequence is a 3' or 5' untranslated region of said mRNA. Preferably said reporter may be a dig label, a fluorophore, a hapten
5 or a radiolabel.

Alternatively said probe may comprise an antibody specific for said vertebrate homologue of said UNC-53 protein.

Preferably the reporter is an antibody conjugated
10 to for example a fluorophore such as fluorescein or alternatively an enzyme such as streptavidin.

As described above, UNC-53 protein of C.elegans has been found to localise to microtubule and particularly to microtubule (+) ends. Therefore,
15 there is provided by a further aspect of the present invention a method of determining whether a compound is an inhibitor or an enhancer of association of the UNC-53 homologue of the invention to microtubules or plus end regions thereof, which method comprises (a)
20 contacting said compound with a transgenic cell, tissue or organism expressing said vertebrate homologue and which protein is operably linked to a reporter molecule (b) screening for the localisation of said reporter molecule as compared to a cell
25 according to step (a) which has not been contacted with said compound.

A compound identifiable by the above method also forms part of the present invention. Such a compound identified as an inhibitor of localisation or
30 association of said vertebrate homologue with microtubules or the plus end region thereof may be used in alleviating the spread of disease inducing cells or metastasis or loss of contact inhibition. Further a compound identified as an enhancer of
35 association of said vertebrate homologue with microtubules or the plus end region thereof may be used in for example promoting neuronal regeneration,

- 22 -

revascularisation or wound healing, or for treating chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease. These compounds may then be included in a pharmaceutical composition, together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

Also provided by the present invention is a kit for determining whether a compound is an inhibitor or an enhancer of association of the vertebrate homologue thereof according to the invention with microtubules or the plus end regions thereof, which kit comprises at least one transgenic cell expressing said UNC-53 vertebrate protein homologue and a reporter molecule or a host or transgenic cell according to the invention and at least one cell of the same cell type for use as a control and means for contacting said compound with one of said at least one transgenic cells. Compounds identified as inhibitors or enhancers or microtubule association described above may advantageously be included in a composition and linked to said vertebrate homologue according to the invention to target the compounds to the microtubules or the plus end regions thereof. Such a composition may also comprise, for example, a suitable transfecting or transformation agent.

According to a further aspect of the invention there is provided a method of targeting a protein to a cell microtubule or the plus end region thereof, which method comprises introducing into a host cell, tissue or organism a transgene comprising a sequence capable of expressing said UNC-53 vertebrate homologue according to the invention, which sequence is operably linked to a sequence encoding said protein to be targeted such that a chimeric protein is expressed and which results in targeting of said protein to said microtubule or a plus end region thereof. An even further aspect of the invention comprises a method of

- 23 -

identifying a molecule which covalently modifies UNC-said vertebrate homologue according to the invention, which method comprises a) contacting either an extract from a cell or cells expressing said vertebrate
5 homologue or a mixture of enzymes comprising candidate UNC-53 modifying enzymes in the presence of an indicator of covalent modification of a protein, b) identifying any covalently modified UNC-53 protein from step a) and c) identifying said molecule involved
10 in said modification step. Such an indicator may be ³²p.

Further provided by the invention is a method of identifying a compound which alleviates or enhances the toxicity of said UNC-53 vertebrate homologue
15 thereof according to the invention, or which alleviates or enhances apoptosis. The method of the former comprises contacting said compound with a transgenic cell, tissue or organism according to the invention and monitoring for the presence of said
20 reporter molecule adjacent said microtubules or the plus end region thereof. In the case of apoptosis the method comprises monitoring the effect of the compound on cell death.

The invention may be more clearly understood from the following examples which are purely exemplary,
25 with reference to the accompanying drawings wherein,

Figure 1(a) is an illustration of the nucleotide sequence encoding the first human homologue of UNC-53 designated Hs-UNC-53/1 and further variants thereof.

30 Figure 1(b) is an illustration of the amino acid sequence of hs-UNC-53/1 encoded by the sequences in Figure 1(a).

Figure 1(c) is an illustration of the nucleotide sequence encoding the second human homologue of UNC-53 protein of C. elegans designated Hs-UNC-53/2 and
35 further variants thereof.

Figure 1(d) is an illustration of the amino acid

- 24 -

sequences of Hs-UNC-53/2 encoded by the sequences in Figure 1(c).

Figure 1(e) is an illustration of a nucleotide sequence encoding the third human homologue of UNC-53 protein according to the invention designated Hs-UNC-53/3, and variants thereof.

Figure 1(f) is an illustration of the amino acid sequences of the Hs-UNC-53/3 encoded by the sequences of Figure 1(e).

Figure 1(g) is an illustration of the nucleotide sequence of a genomic DNA fragment that contains a putative 5' exon of Hs-unc-53/1.

Figure 1(h) is an illustration of the nucleotide sequence AB023155 encoding the protein KIAA0938, a transcript comprising the 3' half of Hs-unc-53/3.

Figure 1(i) is an overview of the *C. elegans* and human UNC-53 proteins as cloned. The 5' truncated variants and a number of the known splice variants have been indicated.

Figure 2 is an alignment of the amino acid sequences of Ce-UNC-53, *Hs-UNC-53/1*, *Hs-UNC-53/2* and *Hs-UNC-53/3*.

Figure 3 is an alignment of the *C. elegans* unc-53 and the predicted amino acid sequence of *C. briggsiae* unc-53.

Figure 4 is a list of ProSite signatures for vertebrate UNC-53s based on the sequence alignment.

Figure 5a is an illustration of expression of the three human UNC-53s as studied by Northern blotting.

Figure 5(b) is an illustration of differential expression of Hs-unc-53/3 in different brain parts.

Figure 6(a) is an illustration of differential splice variant expression of Hs-unc-53/1 using RT-PCR.

Figure 6(b) is an illustration of differential splice expression of Hs-unc-53/2 using RT-PCR.

Figure 6(c) is an illustration of differential expression of Hs-unc-53/3 using RT-PCR.

- 25 -

Figure 6(d) is a sequence confirmation of AB023155 expression in cells other than brain using RT-PCR.

5 Figure 7(a) is an illustration of the cloning of Hs-unc-53/3.

Figure 7(b) is a plasmid map and the nucleotide sequence of the pGI3303 expression vector (C-terminal Hs-unc-53/3 fragment in fusion with GFP).

10 Figure 7(c) is an illustration of the amino acid sequence of GFP: C-terminal Hs-unc-53/3 fragment (insert of pGI3303).

Figure 7(d) is a plasmid map and the nucleotide sequence of the pGI3305 expression vector (full length Hs-unc-53/3 in fusion with GFP).

15 Figure 7(e) is an illustration of the amino acid sequence of GFP : Hs-unc-53/3 (insert of pGI3305).

20 Figure 8 is an illustration of the filipodia and lamellipodia outgrowth of N4 mouse neuroblastoma cells transfected with pGI3303 (F-actin cytoskeleton reorganisation)

Figure 9 is an illustration of the co-localisation of the GFP:Hs-unc-53/3 fusion protein with microtubules in N4 mouse neuroblastoma cells transfected with pGI3305.

25 Figure 11a is an illustration of the homology domains between Hs-unc-53/3 and a gene encoded (partially) by the Drosophila melanogaster BAC clone BACR48M05 (AC005719). Results of a TBLASTN search on the non-redundant database with Hs-unc-53/3 as query.

30 Figure 11b is an illustration of an ORF encoded by the Drosophila melanogaster BAC clone BACR48M05 (AC005719) as predicted by the computer program Fgene.

35 Figure 11c is an illustration of a "BLAST 2 sequences" search result with Hs-unc-53/3 as query and the Fgene predicted UNC53 homology ORF of D. melanogaster BAC clone BACR48M05.

Figure 12 is an illustration of a zebra fish EST

- 26 -

encoding Dr-unc-53/2.

Figure 13 Genemap98 results for Hs-unc-53/2.

Figure 14 is a schematical drawing of the sequence of the exon containing the putative alternative start codon of human Hs-unc-53/1.

Figure 15 is an illustration of the nucleotide sequence of pGI3150 and the amino acid sequence of the eGFP fusion with a C-terminal fragment of Hs-Unc-53/1.

Figure 16 is an alignment of EST clone yk480b6 and Ce-unc-53 demonstrating a novel splice variant of Ce-unc-53.

Figure 17 is a graphical display of the effect of Hs-unc-53/3 GFP chimera transient transfection on the form factor of N4 cells.

DEPOSITED MATERIAL

Plasmids pG13303 and pG13305 were deposited under accession numbers LMBP3936 and LMBP3937 respectively on 28 May 1999 at the Belgian Coordinated Collections of Microorganisms (BCCM) at Laboratorium voor Moleculaire Biologie - Plasmidencollective (LMBP) B-9000 Ghent, Belgium, in accordance with the provisions of the Budapest Treaty of April 28 1977.

Hs-UNC-53/3 is a bona fide UNC-53 (fig. 1; 2; 3)

Blastn and Tblastn EST-database mining using the sequence of the already known animal UNC-53s led to the identification of 3 ESTs suggestive of novel unc-53s (see experimental procedures). By 3'- and 5'-RACE extension using suitable libraries, it was shown that these ESTs identified a novel unc-53 designated Hs-unc-53/3 (Fig. 1 e; f). The publication of the sequence AB023155 (Nagase et al. 1999, DNA Res. 6:63-70) independently confirmed the correctness of the 3'-end of Hs-unc-53/3 as well as the existence of one new

- 27 -

intron that forms the 5'-end of AB023155. Alignments of the *C. elegans* and 3 human UNC-53 sequences (fig. 2) clearly illustrates that the third human homologue of *C. elegans* UNC-53 protein is a bona fide UNC-53 with highest similarity to Hs-UNC-53/2 and in decreasing order to Hs-UNC-53/1 and (*C. elegans* UNC-53) Ce-UNC-53.

Many of the domains of Hs-UNC-53/3 show highest similarity to functional domains of other animal UNC-53s (fig. 2). This critically suggests that Hu-UNC-53/3 most likely has the key functionalities observed for Ce-UNC-53 in a variety of assays including F-actin binding, F-actin reorganisation in cell culture, microtubule and microtubule (+)-end binding in cultured cells, binding of SH3-domain adapters like SEM-5/GRB-2 or other types of binders of proline rich alpha-helices. These results indicate that like Ce-UNC-53, Hs-UNC-53/1, Hs-UNC-53/2, or Hs-UNC-53/3 can be used in a range of biochemical, cellular and animal assays aimed at discovering tissue- or disease-specific modulators of Hs-unc-53 functioning in diagnostic assays.

Further extension of the Unc-53 family (Fig. 11, 12)

Database searches with the three human UNC-53 protein sequences revealed several expressed sequence tags (ESTs) and genomic DNA sequences (BACs) that show significant similarity to human UNC-53.

C. briggsiae

The *C. elegans* genome consortium sequenced the locus of the *C. briggsiae* unc-53 homologous gene. Through gene prediction programs and the cDNA sequence of the *C. elegans* unc-53, prediction

- 28 -

can be made for the *C. briggsiae* protein sequence.
Alignment of the derived *C. briggsiae*
amino acid sequence with the *C. elegans* amino acid
sequence in figure 3 demonstrates the strong homology
5 of both proteins.

D. melanogaster

BAC clone BACR48M05 (AC005719) clearly contains 3
10 different exons with high homology to Hs-unc-53/3
(Figure 11). Using the gene structure prediction
program Fgene [Solovyev et al., 1995, in: Proceedings
of the Third International Conference on Intelligent
Systems for Molecular Biology (eds. Rawling et al.,
15 Cambridge, England, AAAI Press); Solovyev and
Lawrence, 1993, in: Abstracts of the 4th annual keck
symposium. Pittsburgh, 47) it was possible to predict
an ORF encoded by BAC clone BACR48M05 that shows
homology to Hs-unc-53/3 (Figure 11b). However, every
20 *Drosophila* cDNA partially or entirely encoded by BAC
clone BACR48M05 and which contains one or more
sequence blocks as indicated in figure 11a should be
considered as a family member of the UNC-53 family. A
"BLAST 2 SEQUENCE" search indicates that the sequence
25 situated between the three homology blocks that are
indicated in figure 11a is less conserved between
human and *Drosophila* (Figure 11c). The predicted ORF
of the *Drosophila melanogaster* UNC53 gene can be used
to identify new members of the family. The zebrafish
30 EST fc21d06 (AI658309) shows an identity of 84% and a
homology of 92% to Hs-UNC-53/2. It clearly can be
considered as a part of the zebrafish homologue of Hs-
UNC-53/2 (Figure 12). Finally, a whole series of
human ESTs have been placed in public domain
35 databases. To our knowledge, no one has been able to
place these ESTs into contigs that describe a true Hs-
unc-53 to a level presented in this specification.

- 29 -

The presently available unc-53 sequences - expressed or genomic - further underscore that the unc-53 gene family is a true animal gene family in helminths, vertebrates and arthropods, three major classes of the animal kingdom.

Refined UNC-53 family description based on alignment (fig. 4).

The alignment of the three human and the C. elegans UNC-53 sequences enables the more refined definition of conserved regions in UNC-53s. In figure 4 there are compiled a number of proSite signatures for either the four animal or the three human UNC-53s.

Differential expression of Hu-UNC-53/3 by Northern blot (fig. 5).

To determine in which cells and tissues the vertebrate UNC-53s play a role, a northern blot analysis has been performed. As indicated in the experimental section, relevant probes were amplified and used to visualise in which normal human tissues and in which cancer cell lines the three human UNC-53s were expressed.

1. A cancer cell line RNA blots probed with Hs-Unc53/1.

A Northern blot of poly-A+RNA from several cancer cell lines (Melanoma G361, Lung Cancer A549, Colorectal Adenocarcinoma SW480, Burkitt Lymphoma DRajii, Leukemia Molt4, Lymphoblastic Leukemia K562, HeLa S3 and Promyelocytic Leukemia HL60) was probed using the whole insert of pHH3b. No or weak expression was detected in the Burkitt Lymphoma DRajii, the Leukemia Molt4 and the Promyelocytic Leukemia HL60 cell lines. Five different transcripts

- 30 -

are detected in the remaining cancer cell lines:
transcripts 1 and 2 are larger than 9.5kb, transcripts
3 and 4 are 6 to 7 kb and the fifth transcript is
around 6 kb. Transcripts 1 and 2 are present in all
expressing cell lines but at different levels.

Transcripts 3 and 4 are restricted to Melanoma G361,
Lung Cancer A549 (weak) and Colorectal Adenocarcinoma
SW480 and are the predominant transcripts in Melanoma
G361 and Colorectal Adenocarcinoma SW480. Transcript
5 is restricted to Lymphoblastic Leukemia K562 (weak)
and (predominant) in HeLa S3 and is predominant in
HeLa S3.

2. Cancer cell lines RNA blots probed with Hs-
Unc53/2.

A similar set of cancer cell line Northern
blots were probed with a 652bp fragment of EST46037
amplified by using the primers 5'-
aggagatgaagctgacagatatcc and 5'-aaacaccagtgtgagtc. Hs-
Unc53/2 is expressed in Melanoma G361, Colorectal
Adenocarcinoma SW480, Lymphoblastic Leukemia K562 and
HeLa S3. No expression was detected in Lung Cancer
A549, Burkitt Lymphoma DRajii, Leukemia Molt4 and
promyelocytic leukemia HL60. Interestingly only 2
transcript sizes were detected of around 7 kb
expressed in Lymphoblastic Leukemia K562 and HeLa S3
and a transcript of >9.5 kb in Melanoma G361 and
Colorectal Adenocarcinoma SW480 and weakly in HeLa53.
Noteworthy is the very high expression in melanoma
G361.

3. Normal Human tissue probed with Hs-Unc53/1.

A Northern blot of poly-A+RNA from normal
human tissue was probed using the whole insert of
phage HH3b. Expression levels are low in all tissues
with the highest level in heart and placenta, several
fold lower levels in brain and testis, even lower

- 31 -

levels in skeletal muscle, pancreas, thymus, colon, small intestine, ovary and prostate. Expression in peripheral blood leukocyte, lung, liver, kidney, spleen is barely detectable.

5

4. Normal Human tissue probed with Hs-UNC53/2.

A similar set of blots were probed with a 652bp fragment of EST46037 amplified by using the primers 5'aggagatgaagctgacagatatcc and 5'-
10 aaacaccagtgagtcc. Expression levels are low in all tissues with the highest level in kidney, placenta and pancreas, lower levels in heart and lung. Expression is barely detectable or undetectable in skeletal
15 muscle, spleen, thymus, prostate, testis, ovary, small intestine, colon peripheral blood leucocyte, stomach, thyroid, spinal cord, trachea, adrenal gland and bone marrow. Also Hs-unc-53/2 appears to be expressed as different transcripts (figure 5a).

The hs-UNC53/1 and hs-UNC-53/2 homologues are
20 clearly highly regulated genes, showing a strong tissue specificity and, probably, additional mechanisms of regulation (ie differential splicing of different promoters). The different proteins derived from RNA's identified by probe hhl5 presumably share
25 the carboxyterminal nucleotide binding domain. Ce-UNC-53 was shown to be a complex genetic locus and complex transcription unit. The different transcripts are thought to be a mechanism to assure the necessary specificity and functional diversity of this signal
30 transduction pathway, with respect to different signals and receptors, different tissues and different directions of migration. The occurrence of a new transcript or the observed changes in expression levels in the cancer cell line blot suggests a role
35 for hs-UNC-53/3 in the establishment or maintenance of the transformed state of those cells.

- 32 -

Expression pattern of hs-UNC-53/3.

A northern blot of poly-A+RNA from several cancer lines was probed with unique fragments of the three genes from the Hs-unc-53 family. Hs-unc-53/3 has a high expression level in lung carcinoma line A549, where only a moderate expression of hs-unc-53/1 has been detected. Furthermore, moderate expression of Hs-unc-53/3 was also observed in melanoma line G361, where previously, a high expression of hs-UNC-53/1 and hs-UNC-52/2 has been observed. This indicated the involvement of hs-unc53/3 in at least two cancer lines.

In normal human tissues, the expression of hs-unc-53/3 shows a clearly new and previously unobserved expression pattern. This difference of expression of hs-unc-53/3 in relation to its homologues hs-unc53/1 and hs-unc53/2 is important for the allocation of functionality to hs-unc-53/3.

Hs-unc-53/3 is highly expressed in brain, as shown on the Northern blots (figure 5a). In figure 5b it can be seen that Hs-unc-53/3 also is differentially expressed in different parts of the brain. Its homologues are not or weakly expressed in brain. This gives an indication that its function in directionality of cell migration and growth cone steering will be in relation to specific regions or cells of the brain. It is deduced that Hs-unc-53/3 will be an important signal transducer or signal adapter linking signals to neuronal outgrowth, axon guidance, and formation and maintenance of synaptic connections. It seems that the function of Hs-unc-53/3 will be associated with neuron-neuron interactions, neuronal outgrowth, neuron muscle interactions, and post-synaptic signal transduction. Furthermore, Hs-unc-53/3 may be involved in the development of cancer of neuronal origin, like

- 33 -

neuroblastomas, or the development of tumours will have their developmental origin in the brain as some eyes diseases like retinoblastomas.

5 The significance of the high expression of Hs-unc-53/3 in brain tissue can be associated with the high levels of expression which has also been observed in the spinal cord, containing neuronal tissue. Here, neuronal (axon) outgrowth and neuron-neuron connections are of importance. Development of
10 pharmacological tools acting on this pathway may lead to treatments of diseases involved in the growth and movement of neuronal cells, and the regeneration of neuronal connectivity after trauma, or the inhibition of neuronal cancers such as neuroblastomas. Due to
15 its specific expression, inhibitors and/or enhancers specific for Hs-unc-53/3 will have an advantage as a pharmaceutical compound over more general compounds acting on the Hs-unc-53 family of genes and proteins.

20 A second tissue where hs-UNC-53/3 is highly expressed and where (its) other human homologues are not expressed is the spleen. Hs-UNC-53/3 could therefore function as part of the signal transductions pathway involved in the maturation of leukocytes. Malfunction of this pathway may lead to incorrect
25 maturation of the leukocytes and the development of autoimmune diseases such as rheumatoid arthritis and sclerosis. Next to the signalling function in the recognition of the leukocytes, Hs-UNC-53/3 may also play an important role in the induction and/or
30 signalling pathway of the mechanism underlying apoptosis of leukocytes in the spleen. Pharmaceutical methods involving the hs-UNC-53/3 pathway, which may, for example, result in an inhibition and/or enhancement of its expression may lead to treatment of
35 these disorders. Furthermore, hs-UNC-52/2 may have an advantage, as an inhibitor or enhancer specific for hu-unc53/3 which will act in a more specific manner.

- 34 -

The Hu-UNC-53/3 protein is also highly expressed in the ovary, where the two other human homologues are also expressed. Finally moderate to low expression of hs-unc53/3 is observed in heart, placenta, testis, stomach and adrenal gland.

Although the predominant transcripts of Hs-unc-53/3 are > 9 kb, often a smear occurs that ends at with somewhat higher intensity at 5.5 - 6.5 kB. This short transcript may correspond to AB023155.

The Hs-unc53/3 gene is a highly regulated gene, showing strong tissue specificity and additional mechanisms of regulation which have not previously been identified in any of its known homologues. These findings may thus lead to the development of more specific inhibitors or enhancers of hs-UNC-35/3 and or of the Hs-UNC-53/3 pathway. The Northern blot studies indicate that the three human unc-53s are complex transcriptional units with highly regulated tissue specificity and that transcripts of different lengths exist.

Splice variants of human unc-53s

Whilst cloning Hs-unc-53/3, it became apparent that at least three expression variants of Hs-unc-53/3 - most probably alternative splices - exist (fig. 1e, f; lowercase regions). Targeted efforts for the two other human UNC-53s demonstrated that the other human UNC-53s contained variants (fig. 1a, c and e regions).

Splice variants as observed to date appear to be concentrated in specific regions. A first one (starting at position 1252 in fig. 2) - in which the overall amino acid similarity is weak - contains 2 (splice) variants of both Ce-unc-53 and Hs-unc-53/3. In the worm, the presence or absence of these 2 exons in unc-53 regulates the function of the UNC-53 protein in such a way that cells differentially translate

- 35 -

extra-cellular signal gradient as an attractive or repulsive signal. The most 3'-variant of Hs-unc-53/2 roughly covers the 2 Ce-unc-53 variants.

5 The complexity of variation in this zone of Hu-UNC-53 might resemble the situation in the nematode. In Hs-unc-53/3, for example, the region from position 3795 to 4325 (figure 1e) consists of two adjacent blocks (3795 to 4283 and 4286 to 4325 in figure 1e) that can independently be present in or absent from
10 cDNAs from frontal cortex tissue. In contrast, no variants were as yet observed in this zone for Hu-UNC-53/1 or /2.

The second variant in Hs-unc-53/3 (fig. 2) deletes a box (MQLDNRTLPPKKGLR), which is extremely
15 conserved (in bold) among all human unc-53s. This occurrence of this variant could indicate differentially active functional variants of Hu-unc53/3.

A second region in which splice variants were
20 observed contains a major highly conserved domain of unc-53s. Hs-unc-53/1 has a first variant that comprises the most N-terminal portion of this conserved domain (SGSFRD). A second splice variant in Hs-unc-53/1 (AEERMOSE) lies within the highly
25 conserved domain. Another conserved spot for splice variation in human unc-53s has been found (figure 2): Hs-unc-53/1 {VYE}; -/2 {VNE} and -/3 {NSRGSEL}. All these spliced exons are flanked by two conserved charged domains - putative nuclear localisation
30 signals. Given this conservation, we searched for splice variation in C. elegans and found it to exist in the form of an extra exon (ALSVDSQ) (figure 2). Hu-unc-53/3 has another variant (SPLVWPPKKRQNGPVIYKHSR) (fig. 2).

35 The most 3' splice variant in Hs-unc-53/3 has been discovered whilst cloning Hs-unc-53/3 and was shown to be present uniquely in human heart cDNA

- 36 -

libraries.

Single nucleotide polymorphisms

5 Cloning and PCR studies indicated the existence
of a non-silent single nucleotide polymorphism in Hs-
unc-53/1 in position 1232 and in Hs-unc-53/2 in
position 929. This indicated that variations exist in
human unc-53s which - in some cases - may be relevant
10 to the proper functioning of the UNC-53 protein and
hence in disease.

Expression in normal and neoplastic cells by RT- PCR

15 The cloning efforts demonstrated the existence of
splice variants in the human unc-53s and the Northern
blots revealed a range of transcripts for each human
unc-53. The combined data do not explain completely
20 the range of transcripts observed. Therefore, our
understanding of the expression complexity of human
unc-53s may be incomplete and more detailed RT-PCR
studies were performed.

25 One of the obscuring factors could have been that
all studies performed on mRNA or cDNA of whole tissues
which are built of different normal human cell types
that occur in different proportions. For this reason
and because skin was not covered in the Northern blot
studies, a RT-PCR study was set up using cDNA
30 preparations of the different cells in skin normal
human: (1) epidermal keratinocytes, (2) melanocytes,
(3) dermal fibroblasts. In addition, lineage matched
transformed cell lines or tumour cell lines were
included in the study to compare normal versus
35 neoplastic cells. Human umbilical vein endothelial
cells (HUVEC) were taken as a normal human match for
endothelial cell lines.

- 37 -

The RT-PCR study for Hs-unc-53/1 revealed that the most 5'-splice variant is differentially expressed in normal versus neoplastic cells/cell lines. This exon is present in 7/7 keratinocytes, HUVEC and in melanocytes but lacking in HaCat, ECV304, 2/7 melanoma and MCF-7 cells (breast carcinoma).

The RT-PCR study for Hs-unc-53/2 revealed a more surprising picture. The tumourigenic endothelial line ECV304 lacks expression of Hs-unc-53/2, whereas their normal counterpart HUVEC expresses Hs-unc-53/2, suggesting gene deletion or inactivation of expression in ECV304. In epidermal keratinocytes and the lineage matched spontaneously transformed keratinocyte HaCaT and MCF-7 lack expression of the 5'-end of Hs-unc-53/2, but express the 3'-end (starting in or near the microtubule-binding domain). This suggests that like AB023155 for Hs-unc-53/3, also Hs-unc-53/2 can be expressed as a truncated 3'-variant in a cell-specific way. Also splice variation of Hs-unc-53/2 appears to differ in a normal to neoplastic way: the {VNE} exon was shown to be present in all keratinocyte isolates but not in HaCaT and also melanocytes express it, but not 2/7 melanoma or MCF-7. The RT-PCR studies for Hs-unc-53/3 were focussed on demonstrating expression of AB023155 in tissues other than brain. The new exon described was shown to be present in keratinocytes, HUVEC, dermal fibroblasts, melanocytes and their transformed/neoplastic variants, demonstrating its wide expression in tissues in man.

Alternative 5'-start exons

For Hs-unc-53/2 five different start exons have been cloned using RT-PCR, three of which have been confirmed to be present in at least 2 different cDNA libraries (figure 1b, c). Likewise for Hs-unc-53/3 different 5'-exons were found, two of which were

- 38 -

confirmed (figure 1e, f). These 5'-exons most probably indicate that human unc-53s are being expressed via the control of alternative promoters that lie 5' of these different 5'-exons. Also in the
5 nematode has been shown that different (intronic) promoters are driving the expression of 5'-variants of *C. elegans* unc-53.

The Hs-unc-53/1 5'-end

10

Despite considerable efforts, cloning has not lead to the identification of a bona fide 5'-end for Hs-unc-53/1 that comprises an F-actin binding domain, despite the fact that the Northern blots indicate the
15 existence of transcripts > 9.5 kb. Given that both Hs-unc-53/2 and -/3 are expressed as full length and truncated forms, the question can be raised whether Hs-unc-53/1 may not be expressed in a short form as well.

20

cDNA library cloning and 5'-RACE has provided contiguous sequence that ends at a position that matches with a domain in *C. elegans* un-53, where an alternative start position lies. Based on this
25 argument, Hs-unc-53/1 could be a functional equivalent in man of this transcript in nematode.

25

To further trace the "longer" variants of Hs-unc-53/1, genomic BAC DNA sequencing has been performed. In figure 1g is shown sequence of a 4984 fragment from BAC 585E09. It comprises sequence 5' of the presently
30 known cDNA of Hs-unc-53/1. To the qualified as well as by means of two groups of gene structure prediction computer programs, different but comparable exons in the 4984 bp genomic sequence fragment can be predicted (figure 14). The programs GENSCAN, HEXON and MZEF all
35 predict an exon between bp 1089 and bp 1880. The end of this predicted exon (bp 1880) is confirmed by the cDNA sequence. Therefore this predictions has a big

- 39 -

change to indicate the correct exon length. The programs GRAIL, GENEFINDER and HMMGENE all predict an exon between bp 1123 and bp 2031. None of the predicted exons contains an in frame stop codon 5' of the alternative start codon. Consequently, it is possible that there exist unidentified exons 5' of the exon containing the alternative start codon.

The present picture critically suggests that both nematode and human unc-53s appear to be complex transcriptional units. Moreover, the fact that some of the most complex splice variants map to similar regions in the UNC-53 proteins points to evolutionary conserved functional variants of UNC-53s e.g. with regard to the cells directional migration towards or away from a signal source. In contrast, some of the variants in the human UNC-53s are located in highly conserved domains; these (and other) variants may create discrete - yet undiscovered - functionally different UNC-53 proteins transcribed from one of the unc-53 genes.

The fact that two and maybe three human unc-53s exist as full size and a truncated forms with cell-specific expression, that series of alternative 5'-start exons exist eventually controlled by different promoters that some forms of splice variation are conserved from nematode to man, all indicate that the expression of unc-53s is of very high complexity and that some of the biological functions of UNC-53 proteins are extremely conserved.

On the other hand, the differential expression in Northern blots, the splice variation difference between normal and lineage-matched neoplastic cells and the non-silent single nucleotide changes in two of the three human unc-53s, all indicate how important a wide range of diagnostic assays can be to understand in depth the role in disease of human unc-53s.

- 40 -

**Chromosomal localization of Hs-unc-53/2 by
Genemap98 (Fig. 13 and 1(c))**

The EST clones AA918601, AI248585, AA115014 and
5 AA115015 are clearly homologous to the 3'-UTR of Hs-
Unc-53/2 cDNA (Figure 1(c)). Although, AA115014
(describing the same EST as AA115015) contains an
alternative splice variant of the Hs-Unc53/2 gene in
the 3'UTR. A survey with ESTs AA918601, AI248585,
10 AA115014 or AA115015 as query in the genemap98
database (release November 1998) revealed that the Hs-
Unc53/2 gene is located at chromosome 11
([http://www.ncbi.nlm.nih.gov/genemap98/loc.cgi?ID=2122](http://www.ncbi.nlm.nih.gov/genemap98/loc.cgi?ID=21224)
4). The STS which is used for chromosomal
15 localization and which is situated in the 3'UTR of the
Hs-Unc53/2 gene is referred to as SHGC-33456 (dbSTS
Id: 41891, Genbank Acc: G28036, Genbank gi: 1396755)
(Figure 13a). The STS was localized by analysis on
the NIGMS human/rodent somatic cell hybrid panel
20 (dbSTS Id: 41891). The Radiation hybrid results are
summarized in Figure 13b. Together these data imply
that every disease or phenotype connected to SHGC-
33456 is due to the Hs-Unc-53/2 gene.

25 **Functional Characterisation of Hs-unc-53/3**

**F-actin reorganisation and microtubule binding of
Hs-unc-53/3**

30 Based on its structural features, Hs-unc-53/3 can
be classified as a bona fide human unc-53. To further
understand its function and in anticipation of
developing pharmacological compound screening assays,
Hs-unc-53/3 has been physically cloned following the
35 method described in the experimental section and shown
in figure 7a. The derived Hs-unc-53/3 clones
comprising full length (A to L and the 3'-half (G to

- 41 -

L) of Hs-unc-53/3 were further engineered to form a chimera with green fluorescent protein and cloned into expression vectors appropriate for transfection of eukaryotic cells. The nucleic acid and amino acid sequences of these constructs are shown in figure 7b-e. The constructs were transfected into cells and scored for their effects on the F-actin cytoskeleton and binding to microtubules of mouse neuroblastoma cells N4; functions known for nematode unc-53 and human unc-53/1.

The N4 cell transfected with a GFP fusion to the 3'-half of Hs-unc-53/3 (pGI3303, fig. 7b) showed pronounced filopodia and lamellipodia outgrowth, which is associated with reorganization of the F-actin cytoskeleton (Figure 8). This observation demonstrates that like nematode unc-53 and human unc-53/1, the F-actin binding domain is not required for inducing reorganization of the F-actin cytoskeleton of N4 cells. In addition, the pGI3303 encoded fusion protein does not co-localize with microtubuli but localizes to the cytoplasm of N4 cells indicating that an important domain for microtubuli association is missing in this C-terminal fragment of Hs-unc-53/3. In the alignment figure 2 can be seen that the C-terminal half of Hs-unc-53/3 (approximate KIAA0938) does not comprise the conserved microtubule binding domain.

In contrast, the N4 cells that expressed low to medium levels of the GFP fusion to full length Hs-unc-53/3 (pGI3305, Fig. 7d) displayed a co-localization of the GFP fusion protein with microtubules (Figure 9). Even the centrosomes could clearly be detected in some transfected cells. Cells expressing very low amounts of the fusion protein displayed specific microtubule (+)-end binding (Figure 9). The morphology of the pGI3305 transfected N4 cells does not clearly differ from the control transfected cells although there is a

- 42 -

tendency towards rounding up of the pGI3305 transfected cells and filopodia outgrowth.

Validation of functional assays as compound screens

5

R74288 has previously been shown to be an inhibitor of nematode function in *C. elegans* (WO96/38555), an activity that has been confirmed in
10 *Ce-unc-53* transfected N4 cells, where only the transgene-induced effect was inhibited by R74288. In order to confirm compound R74288s activity in a full mammalian system, a stable transfection of plasmid pGI3150 was performed in the N4 neuroblastoma cell
15 line with the lipofectamin procedure (Gibco BRL). pGI3150 expresses an eGFP protein in fusion with the C-terminal end of Hs-unc-53/1 (see Figure 15a). After two weeks of G418 selection, 20 clones with stable integration of the pGI3150 plasmid were selected and
20 isolated. These clones were tested for GFP expression by fluorescence microscopy and by Western blotting with an anti-GFP antibody (table 1). The lamellipodia outgrowth phenotype was checked visually (See Figure 15b). Compound R74288 was tested on four random
25 selected pGI3150 stably transfected clones: 8.1, 8.2, 8.3 and 10.1 and on a pool of pEGFPC1 stable transfected N4 control cells. Clones 8.2 and 10.1 displayed less lamellipodia outgrowth than clones 8.1 and 8.3. Compounds and solvents were added to the
30 stably transfected cells (10^5 M in DMSO). After 24 hrs of incubation, two persons independently scored the effect of the treatments on the cells. As shown in table 1, both persons noticed an effect compound 2 on clones 8.2 and 10.1 with a weak transgene-induced
35 lamellipodia phenotype. This effect consisted of a more flat morphology of the treated versus untreated cells. Compound 2 was R74288.

Table 1. Effect of compounds on lamellipodia formation

Clone	Compound 1	Compound 2	Compound 3	Compound 4	GFP fluo	GFP Western	Phenotype
5 8.1	0	0	0	toxic	+	+	+
8.2	0	+	0	toxic	++	+++	+/-
8.3	0	0	0	toxic	++	++	++
10.1	0	+	0	toxic	+/-	+	+/-
10 GFP pool	0	0	0	toxic			-

Automated compound screening by measuring cell morphology

15 Compound screening assays must have a sufficiently high throughput to be relevant to drug discovery. To achieve this goal, we automated the procedure of measuring the morphological changes induced in cells following transient transfection with

20 full length or 3'-half of Hs-unc-53/3 GFP chimeras. The cell culture, transfection, fluorescence staining and microscopy procedures are performed within a 96-well plate (all-in-one). The fluorescent staining method comprises a triple fluorescent labeling

25 procedure (1) for cell nucleic using DNA double helix intercalating dyes such as Hoechst 33342 or DAPI, (2) for transfection efficiency and expression level of the chimeric protein using GFP fluorescence and (3) for the F-actin cytoskeleton using fluorescently

30 labeled phalloidin, a microfilament dye.

These three different fluorescent images are collected using an motorised stage plus stage driver and a frame grabber that produces seamless composite images of the cells in the well. The software

35 programs to drive this operation are known in public domain as "SCIL" (University of Amsterdam). The seamless images are then superimposed using

- 44 -

pseudocolour for the operator to inspect the quality of the culture. In addition, the SCIL program was compiled in such a way that it: (1) identifies cells by means of their nucleus, (2) measures the GFP
5 fluorescence intensity, (3) delineates the area of the F-actin (phalloidin)staining surrounding a nucleus and (4) calculates a range of parameters objectively representing the features of the F-actin staining pattern of each individual cell. One example of such
10 a parameter is called the "form factor". It is an arbitrary value that reflects the dendricity of a cell. It is derived by calculating (A) the true circumference of a cell's F-actin staining area as seen in the image and (B) the area of the F-actin
15 staining of that given cell. The ratio $4\pi \text{Pix}(B)^2 =$ the form factor. For a rounded cell, the form factor approximates 1 whereas, for a cell with increased filopodia and lamellipodia outgrowth, the true circumference will be much larger than that of a
20 circle and as a result, the form factor $\ll 1$.

In experiments it was shown that transiently transfected N4 cell populations indeed displayed a different form factor versus control cells. Both the median and average form factor for a cell population
25 in a well were reduced following transfection with the 3'-half of Hs-unc-53/3. More in particular, there was a significant decrease in the number of cells in a transfected culture that displayed the minimal form factor, suggesting that the Hs-UNC-53/3 transgene
30 induced round cells in particular to become more dendritic (figure 16).

Chromosomal localisation of Hs-unc-53/3 by FISH indicative for a role disease

35

With FISH technology using a unique fragment of hs-unc-53/3 we are able to localize the hs-unc53/3

- 45 -

gene on chromosome 12q21.1. Chromosome 12q21.1 is a region shown to be involved in autosomal dominant, cornea plana and closed angle glaucoma (Sigler-Villanueva et al., Ophthalmic Genetics 18:55-62, 1997). This indicates that hs-UNC-53/3 protein may be involved in eye development and thus eye diseases, such as retinoblastomas. Neuroblastoma cell line NPG and liposarcoma line WDLPS and other sarcoma lines have amplifications in this region. The neuroblastoma amplification seems to be located more distal (12q24) while the liposarcoma line is located at 12q21 (Van Royal et al., Cancer Genetics and Cytogenetics 82:151-4, 1995). Three loci related to Darier's disease, an autosomal dominant genodermatosis disease characterized by epidermal acantholysis and dyskeratosis have been mapped in region 12q21-q24 (Wright et al., Journal of Investigative Dermatology 103:665-8). 12q21 is also known to be a fragile site associated with the pathogenesis of non-Hodgkin's Lymphoma (Chary-Reddy et al., Cancer Letter 86:111-7 1994). Duplications related to nephroblastoma tumorigenesis were commonly found in the 12q21-q23 region (Austruy et al., Genes Chromosomes Cancer 14:285-294, 1995). In a girl with mental retardation, a conclusive disorder and clinical findings resembling cerebral palsy, positioning of segments from other autosomes adjacent to the band 12q21 were found (Biederman et al., Ann Genet 19:257-260, 1976). Cytogenetic analysis for myeloid leukemia showed a complex karyotype with chromosomal breakpoints at 12q21 (Weinstein et al., Cancer Genet Cytogenet 48:75-81, 1990). Finally, analysis of complex chromosomal rearrangements in malformed children and from spontaneous abortions showed specific breakpoints at site 12q21 Gorski et al., Am J Med Genet 29:247-261, 1997). Most of these diseases have been shown to be involved with cell movement, aberrant development, or

- 46 -

cell-cell contact and neuronal tissue or neuronal development.

Confirmation of FISH with Radiation hybrid panels

5

To confirm and refine the chromosomal localisation of the human unc-53s an alternative method for FISH has been used. Radiation hybrid (RH) mapping is a somatic cell hybrid technique that was developed to construct high-resolution, contiguous maps of mammalian chromosomes. RH mapping provides a method for ordering DNA markers spanning millions of base pairs of DNA at a resolution to easily obtained by other mapping methods. Some of the advantages of RH mapping are (1) distance estimated by this method is directly proportional to physical distance, (2) nonpolymorphic DNA markers, that can not be used for meiotic mapping, can be used for this method, and (3) a high resolution map that is not easily made by other methods can be obtained.

20

The results of FISH and RH mapping for the three human unc-53s are summarised in table AA. By using publicly available databases (see experimental section) one can derive information on the correlation between FISH and RH mapping. RH Mapping was shown in this way to confirm the FISH data for the three unc-53s.

25

Table 2. RH Mapping Primers and Results

Unc-53	FOR Primer	REV primer	PCR Results	Marker*	FISH
5 Hs-UNC-53/1 (BAC585E9)	5' TGTGGGT GAGGAATGC TGAC	5' CAGAGCTT GCTCTAGAGG AC	51, 62, 66	SHGC-30236	1q31-32
Hs-unc-53/1 (BAC585E9)	5' CCTGCCC AACATAGCA AGAC	5' CCATCTAC AATGAGCCAG AC	51, 62, 66	SHGC-30236	1q31-32
10 Hs-unc-53/2 G411	5' CTGCCTC CCTTTGCTG TGTTGCATG	5' CTGAGCAG AGTGAAGCCA GAGTTGG	8, 28, 29, 43, 44, 51, 59, 66, 70, 77, 83	AFM022th2	11p15.t
Hs-unc-53/2, F4.1.2	5' TCATGTA TTCCCCACA GACAAGCC	5' CATTGTGT CTTGATACTT TGGGGTGC	8, 28, 44, 51, 59, 65, 83	SHGC-31021	11p15.1
Hs-unc-53/2, D4.1.1	5' GAGGATT TTATTTCTG GGAAATGGA ATCGG	5' TGATCTTC CACTCCGTGG ATAACT	8, 27, 28, 29, 43, 44, 51, 59, 65, 70, 83	AFM022th2	11p15.1
15 Hs-unc-53/2, J4.1.4	5' AAAGCCC AAGCCCCGG GAGAAGATG	5' AACCCGTT TTCCACCGAG CCGCTC	8, 27, 28, 43, 44, 51, 59, 66, 70, 83	AFM022th2	11p15.1
Hs-unc-53/3, A215	5' ACTTGCT GAAACAGAG AGCTCCATG	5' CTTGCTGT CTTCTTTCTC CTTGGC	1, 48, 50, 51, 59, 65, 66, 73, 74, 76, 78	SHGC-17536	12q21.1
Hs-unc-53/3, A211	5' TGATCTT CTAGCGTGT GACTCACTG	5' ATCATTCC TTGGAGT	1, 48, 50, 51, 59, 73, 76, 78	SHGC-17536	12q21.1

20 (*) list not exhaustive

Also sequence information available in public domain can help refine the positioning of the unc-53 genes, like in the following example. The EST clones
 25 AA918601, AI248585, AA115014 and AA115015 are clearly homologous to Hs-Unc53/2 cDNA. Although, AA115014 (describing the same EST as AA115015) contains an alternative splicevariant of the Hs-Unc53/2 gene in the 3'UTR. A survey with ESTs AA918601, AI248585,
 30 AA115014 or AA115015 as query in the genemap98 database (release November 1998) revealed that the Hu-

- 48 -

unc53/2 gene is located at chromosome 11
(<http://www.ncbi.nlm.nih.gov/genemap98/loc.cgi?ID=21224>). The STS which is used for chromosomal
localization and which is situated in the 3'UTR of the
5 Hs-Unc53/2 gene is referred to as SHGC-33456 (dbSTS
id: 41891, Genbank Acc: G28036, Genbank gi: 1396755)
(Figure 13). The STS was localized by analysis on the
NIGMS human/rodent somatic cell hybrid panel (dbSTS
id: 41891). The radiation hybrid results are
10 summarized in Figure 13. Together these data imply
that diseases or phenotypes connected to SHGC-33456 is
due to the Hs-Unc53/2 gene.

EXPERIMENTAL PROCEDURES

15

Cloning & sequencing of Hs-unc-53/3

Hs-unc53/3 has been cloned starting from a series
of ESTs that were similar but not identical to Hs-unc-
20 53/1 or -/2. The ESTs were:

1. WashU-Merck EST 767735.

Transformed cells carrying the EST 767735
25 sequence were ordered from Research Genetics. Plasmid
DNA was isolated using standard protocols (Qiagen
plasmid DNA isolation kit), the sequence of the insert
was determined.

- 30 2. ATCC cDNA clones 86459.

Transformed cells carrying the cDNA clone
86459 sequence were ordered from ATCC. Plasmid DNA
was isolated using standard protocols (Qiagen plasmid
35 DNA isolation kit), the sequence of the insert was
determined.

- 49 -

3. Genethon cDNA clone c09a03 from the Geneexpress cDNA program.

5 Transformed cells carrying the cDNA clone c09a03 sequence were ordered from Genethon. Plasmid DNA was isolated using standard protocols (Qiagen plasmid DNA isolation kit), the sequence of the insert was determined.

10 These ESTs were extended to form one ORF as follows:

1. 5' extension of EST 767735 by RACE (Rapid Amplification of cDNA Ends).

15

Marathon-Ready cDNAs (Clontech) are premade "libraries" of adaptor-ligated double-stranded cDNA ready for use as templates in RACE experiments. Five ml Marathon-Ready cDNA was used as template in a regular 50 ml RACE. The RACE mixture contained 1 x KlenTaq PCR buffer. 0.2 mM of each dNTP, 1 x advantage KlenTaq polymerase mix (Clontech), 0.15 mM AP1 adaptor primer and 0.15 mM RACE gene specific primer. The amplification conditions were as follows:

20 94°C for 30 s and 68 °C for 4 min. One-hundred-fold diluted RACE product was used as a template in a nested PCR with AP2 adaptor and gene specific nested PCR primers. Specific nested PCR fragments were cloned into pCR2 (TA cloning kit, Invitrogen) and the sequences of the inserts were determined. Gene-specific primer (hh3UNC53 97101702):

25 5' ACCATTTACACCTGAAGACGATTGAGGTCC; nested gene-specific primer (hh3UNC53 97101701)

30 5' CTCCTATTTAATTAGAGGCTCCCTGGACC Marathon cDNA library: human placenta, human heart, human chronic myelogenous leukemia, human colorectal adenocarcinoma.

35

- 50 -

2. 3' extension of EST 767735 by RACE.

Method as described previously. Gene
specific primer (hh3UNC53 97102702)

5 5'CAATCGTCTTCAGGTGTAAATGGTAACGTG; nested gene specific
primer (hh3UNC53 97102703)

5'GAATGTCAAACACAGTGCCACCTCCACC Marathon cDNA library:
human placenta, human heart, human HeLa, human
melanoma.

10

3. 3' extension of cDNA clone c09a03 by RACE.

Method as described previously, gene-
specific primer (hh3UNC53 98020401)

15 5'AGGGAGCACTGAATGGTCCAGACCATCCTC; nested gene-specific
primer (hh3UNC53 98020402)

5'GCATCAGAAGACAGCATTCCTCTGAAAGTG Marathon cDNA
library: human placenta, human heart, human HeLa,
human melanoma, human colorectal adenocarcinoma, human
20 chronic myelogenous leukemia.

4. 5' extension of cDNA clone 86459 by RACE

(1).

25 Method as described previously gene-specific
primer (hh3UNC53 98020403)

5'TTCAATTTCTATCTCTATGAGTTTTCTTCG; nested gene-specific
primer (hh3UNC53 98020404)

5'GCAGCTCTAGATTTGGTGATGAAGAACTC Marathon cDNA
30 library: human placenta, human heart, human HeLa,
human melanoma. Overlapping sequences were assembled
in a single contiguous sequence.

5. 5' extension cDNA clone 86459 by RACE (2).

35

Method as described previously gene-specific
primer (hh3UNC53 98022502)

- 51 -

5' TCAGAATGTGATGAAGGAGGCTTGGTGGAC; nested gene-specific primer (hh3UNC53 98022501)

5' GGATGCCGGAAGGGATGAATCAGTAAGC Marathon cDNA library: human placenta, human heart, human HeLa, human melanoma, human colorectal adenocarcinoma, human chronic myelogenous leukemia.

Validating variants at 5' end of the cDNA sequence

10

In the final 5' RACE experiment, 2 variants have been found whose sequence diverge upstream from the IYTDWAN protein sequence (position 289 in figure 1e or position 82 in figure 1f). By using primers
15 ATTTACACTGACTGGGCCAAC and ATAATCTGGATGATTTCTGCTAGGAGT on cDNA clones a Hs-unc-53/3 specific PCR product was obtained that was radiolabeled using the random primed DNA labeling kit (Roche Molecular Biochemicals) and hybridized to human DNA BAC filters (Research
20 Genetics). Both primers are located near the IYTDWAN box. Four BACs turned out positive (415J11; 464C17, 525C02 and 537B02). DNA sequencing of the region upstream from the IYTDWAN protein sequence directly on these BACs showed that this region was preceded by a
25 putative intronic sequence as evidenced by the multiple stop codons in the reading frame and by the consensus AG intron acceptor sequence. For sequencing purposes, BAC DNA was prepared according to a modified Qiagen plasmid DNA procedure.

30

A primer pair was designed specifically to amplify the 5' end of the variant shown in full in figure 1e (primers ACTTGCTGAAACAGAGAGCTCCATG and CTTGCTGTCTTCTTTCTCCTTGGC). PCR with these primers on BAC DNA showed the presence of the genomic sequence
35 encoding this variant in 3 out of the 4 BACs (not present in BAC 415J11).

- 52 -

BACs containing the genomic sequence encoding the other 5' end variant of Hs-unc-53/3 as shown as the variant in figure 1e were identified by hybridizing the Research Genetics human DNA GAC filters with
5 primer TGATCTTCTAGCGTGTGACTCACTG, radioactively labeled using gamma-P32-ATP and polynucleotide kinase. Positive BACs were 404F14, 450K18 and 764L15.

Sequencing directly on the respective BACs in the 3' direction from within the 2 alternative 5' exons and comparison of the genomic DNA sequence with the previously determined cDNA sequence identified the GT intron donor site. Joining of the genomic sequences from both 5' exons and the IYTDWAN encoding sequence
15 after removal of the predicted intronic sequence restored for both variants the sequence of the 5' RACE experiment without affecting the translation of the Open Reading Frame.

20 Cloning of Hs-unc-53/3 constructs

With the aim of cloning the full-length Open Reading Frame of Hs-unc-53/3, primer pairs were selected such that the ORF could be amplified in 6
25 overlapping fragments ranging in size from 1 to 2 kbp. Overlaps between the fragments were chosen such that they contain an endonuclease restriction enzyme recognition site suitable for cloning the full-length gen. For the 5' fragment, the downstream oriented
30 primer was chosen to contain the first putative start codon (ATG) in variant 1 (the one shown in full in figure 1e). PCR conditions using the Expand High Fidelity PCR system (Roche Molecular Biochemicals) for all of the fragments were as follows. Initial
35 denaturation for 5' at 95°C; 30 cycles of denaturation at 95°C for 45", primer annealing at 55°C for 45" and extension at 72°C for 1' (3' for primer combination

- 53 -

A+B); followed by an additional incubation for 7' at 72°C and storage at 4°C. PCRs were run on PE Biosystems 9700 PCR machines.

5

Primer pairs used for cloning Hs-unc-53/3 fragments

	#	Size (bp)	Primer	Sequence
10	A-B	2229	A	TCAGCTCGAGCATATGCCTGTTCTTGGGGTTGC
			B	GGGGTGGGTGCGACTTGTCAAGTGG
	C-D	847	C	ATGGAAGGACCATAACCAACTTGAC
			D	CTTGTTCCAGCTTTCTGCCTAGATG
15	E-F	781	E	CAGGTTCTGGAGAAGAGGCATGTC
			F	GGTGAGGCAATATCTGGATACTTGG
	G-H	1291	G	AGGCAGCCAGGATCCAAGTATCCAG
			H	TGCGAAGATCTTTTGGGAGGATGGTC
20	I-J	1022	I	AACCATTGAAATGCTGAAGGCTCAG
			J	GGTTATGGGATCTAATTAAGTCTCC
	K-L	1255	K	CACTAGCCTTGGTCTGAGCTCTGAC
			L	TCACCCTCTAGAGGGTAGATTCAAG

Primer A contains restriction sites (XhoI and nheI) suitable for final subcloning in an eukaryotic expression vector (pEGFPc3) and in a yeast-two-hybrid vector (pAS2-1), respectively.

PCR products were analyzed by agarose gel electrophoresis and were visualized by ethidium bromide staining. Splice variants as mentioned in figure 1e were observed as multiple bands on agarose gels. Single band PCR products were purified with the Qiaquick PCR purification kit, whereas multiple band PCR products were cut out from gel as individual bands and purified using the Qiaquick gel extraction kit. PCR products were cloned in pCR2.1 according to the suppliers protocol (Invitrogen). For each fragment, multiple clones were picked from selective LB agar plates and grown overnight under antibiotic selection pressure for DNA preparation either on the biorot 9600

- 54 -

(Qiagen), or manually on anion exchange columns (Qiagen tip 20 or tip 100). Insert sequences were determined using the Bigdye terminator ready reaction cycle sequencing kit (PE Biosystems). Individual sequencing reactions for each clone were assembled in single sequence contigs using the Sequencher software package (GeneCodes). Sequences were compared to the previously determined consensus sequence using the SeqEd software package from PE Biosystems. For each fragment a clone was selected containing the correct sequence and the splice variant of interest. For the I-J fragment, a clone was selected that missed the hart specific 22 amino acid splice variant (figure 1f). In the K-L fragment clone, a SfiI-SacII linker was cloned in the BamHI site of the pCR2.1 multiple cloning site to facilitate subcloning of the full-length gene into the yeast-two-hybrid vector (pAS2-1) and the eukaryotic expression vector (pEGFPc3), respectively.

The overall cloning strategy of the full-length gene is visualized in figure 7a. 7a1 illustrates the overlapping PCR fragments and the nomenclature of fragments and primer pairs. 7a2 illustrates the assembly of the 3' half of the gene in pCR2.1. Internal BamHI (I-J fragment) and XhoI (K-L fragment) sites as well as restriction sites from the multiple cloning site of pCR2.1 (as shown in the figure) were removed by site-directed mutagenesis (SDM) using the Quickchange Site-Directed mutagenesis kit (stratagene). The NotI-EcoRI G-H fragment and the EcoRI-NheI I-Jd22 (d22 indicating that the 22 amino acid splice variant is absent) were directionally cloned in the NotI and NheI sites of the K-L fragment clone. Multiple clones were picked and verified by DNA sequencing. 7a3 illustrates the assembly of the 5' half. Internal XhoI (C-D fragment) and SfiI and XhoI (E-F fragment) sites were removed by SDM.

- 55 -

Inserts were cut out from the vectors by restriction digestion with the appropriate restriction enzymes (XhoI+SaliI; SaliI+NarI and NarI+BamHI, respectively) and purified from gel after agarose gel electrophoresis. The 3 fragments were ligated together, re-cut with XhoI and BamHI and separated on gel. The band of the expected size was cut out of gel, purified and cloned in front of the 3' half, opened by digestion with XhoI and BamHI (figure 7a4). Multiple clones were picked and verified by sequencing.

Figure 7a illustrates the modular nature of the cloning project. For all the possible combinations of splice variation within the building block fragments, one representative clone is available. In view of functional analysis, building blocks can be exchanged easily by standard technology, either in the pCR2.1 construct or in the final eukaryotic expression or yeast-two-hybrid construct.

Construct of Hs-unc-53/3 GFP chimeras

The construction of the mammalian expression vectors pGI3303 and pGI3305 is explained in the legends of figure 7a, 7b and 7d. pGI3303 can be used to over-express in mammalian cells or animals a fusion protein between eGFP and 1128 AA C-terminal fragment of Hs-unc-53/3 (Fig 7c). pGI3305 can be used to overexpress in mammalian cells or animals a fusion protein between eGFP and the 2363 AA full length Hu-unc-53/3 (fig 7d). The Hs-unc-53/3 cDNA in pGI3303 as well as in pGI3305 contains silent mutations that introduce or remove specific restriction sites in order to be able to easily subclone different types of alternative splice variants in these vectors.

- 56 -

Genomic DNA sequencing (BAC 585E09)

Using the primers AGGACCCTATGCGGAGGTCAAGCCGC and TGGGTTGGCATCATCGCTGTCGTAGC, a PCR specific for Hs-unc-53/1 was developed. PCR products were radiolabeled using the Random Prime DNA labeling kit (Roche Molecular Biochemicals) and hybridized on the human genomic DNA BAC filters (Research Genetics). Positive signals were obtained for BAC clones 366H21, 483L14, 471J09 and 585E09. BAC DNA was isolated from E. coli genomic clone 585E09 according to a modified Qiagen plasmid DNA preparation procedure. A shotgun library of 1920 clones was constructed at GATC (Konstanz, Germany). BAC DNA was prepared, nebulized and subcloned after end-repairing in the sequence vector pTZ19R. At JRF, DNA was prepared on the Biorobot 9600 (Qiagen) from 1440 clones. End sequencing reactions with M13 forward (TGTAACGACGGCCAGT) and reverse (CAGGAACAGCTATGACC) primer were done on 768 clones. 672 additional clones were sequenced with M13 only. 5 μ l DNA was used in 15 μ l final reaction volume using the BigDye Terminator Ready Reaction sequencing kit. Sequencing reactions were run on MJ Research PTC200 PCR machines. Reaction products were run and analysed on PE ABI 377 DNA sequencers. All sequencing results were imported in the Sequencher (GeneCodes) software package. Contaminating vector sequences and trailing sequences of low quality were trimmed. Individual sequences were assembled in contigs with standard software settings. A great number of contigs were constructed ranging from below 500 bp to over 10 kbp. Singletons are also still present. By looking for strings of known sequence, a contig was found containing the known and reliable 5' end of hUNC53h1 and extending this sequence in 5' direction. This sequence and its relevant features are described in figure 1g and its legend.

- 57 -

Northern blotting

A Human multiple tissue Northern (MTN-1, Clontech) containing in each lane 2 mg of poly A + RNA from eight different human tissues (heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas) and a MTN-II human multiple tissue Northern, containing in each lane 2 mg of poly A + RNA from spleen, thymus, prostate, testis, ovary, small intestine, colon and peripheral leukocyte, were hybridized according to the manufacturer's instructions and washed out in 0.1xSSC:0.2% SDS at 55°C. Also from Clontech, a poly A + RNA blot from human cancer cell lines (melanoma G361, lung carcinoma A549, colorectal adenocarcinoma SW480, Burkitt's lymphoma Raji Leukemia Molt 4, lymphoblastic leukemia K562, HeLa S3 and promyelocytic leukemia HL60) was tested.

Cancer cell lines RNA blots probed with Hs-unc-53/3

A set of cancer cell line Northern blots were probed with a 665 bp fragment of Hs-unc-53/3 amplified by using the primers 5'AGGAATTAAATTAACGGATATTCGG and 5'AAACTGTCCAACTATTTCTTCTACC. HU-unc-53/3 is expressed in Melanoma G361 and lung carcinoma A549, transcripts sizes were detected of >0.5 kb. No expression was detected in promyelocytic leukemia HL-60 HeLa cell S3, chronic myelogenous leukemia K-562, leukemia MOLT-4, Burkitt's lymphoma Raji and colorectal adenocarcinoma SW480.

Normal human tissue RNA blots probed with Hs-unc-53/3

A set of normal human tissue Northern blots were

- 58 -

probed with a 665 bp fragment of Hs-unc-53/3 amplified by using the primers 5' AGGAATTAAATTAACGGATATTCGG and 5' AAAACTGTCCAACTATTTTCTTCTACC. High expression levels were detected in brain, spleen, ovary and spinal cord, lower levels in heart, placenta, testis, stomach, and adrenal gland. Transcripts sizes were \geq 9.5 kb.

FISH

10

Hs-UNC-53/3 is localised to chromosome 12q21.1

Slides preparation:

15

Lymphocytes isolated from human blood were cultured in α -minimal essential medium (MEM) supplemented with 10% foetal calf serum and phytohaemagglutinin (PHA) at 37°C for 68-72 hr. The lymphocyte cultures were treated with BrdU (0.18mg/ml Sigma) to synchronise the cell population. The synchronised cells were washed three times with serum-free medium to release the block and recultured at 37°C for 6 hr in a α -MEM with thymidine (2.5 μ g/ml: Sigma). Cells were harvested and slides were made by using standard procedures including hypotonic treatment fix and air-dry.

20

25

In situ hybridisation and FISH detection:

30

A cDNA probe was biotinylated with dATP using the BRL BioNick labelling kit (15°C, 1 hr) Heng et al, 1992). The procedure for FISH detection was performed according to Heng et al., 1992 & Heng and Tsui, 1993. Heng et al.: Proc Natl Acad Sci USA 89: 9509-9513 (1992). Heng et al. Chromosoma 102: 325-332 (1993). Briefly, slides were baked at 55°C for 1 hour. After RNase treatment, the slides were denatured in 70%

35

- 59 -

formamide in 2xSSC for 2 min. at 70°C followed by dehydrated with ethanol. Probes were denatured at 75°C for 5 min. in a hybridisation mix consisting of 50% formamide and 10% dextran sulphate. Probes were loaded on the denatured chromosomal slides. After over night hybridisation, slides were washed and detected as well as amplified. FISH signals and the DAPI banding pattern were recorded separately by taking photographs, and the assignment of the FISH mapping data with chromosomal bands was achieved by superimposing FISH signals with DAPI banded chromosomes (Heng et al, 1993).

Results

Under the condition used the hybridisation efficiency was approximately 67% for this probe (among 100 checked mitotic figures, 67 of them showed signals on one pair of the chromosomes). Since the DAPI banding was used to identify the specific chromosome, the assignment between signal from probe and the long arm of chromosome 12 was obtained. The detailed position was further determined in the diagram based on the summary from 10 photos.

Radiation Hybrid Mapping

Radiation hybrid analysis is a PCR technique and the panels of radiation hybrid DNA are provided at a concentration of 25 ng/ μ l in TE buffer suitable for these reactions. Typically, 25 ng of DNA is used in a 10 μ l PCR reaction.

Some of the radiation hybrid panels are supported by an e-mail server which can assist you in the chromosome localization of markers. A server for the chromosome localization of markers using the Stanford G3 and Stanford TNG panels is available at <http://www->

- 60 -

shgc.stanford.edu. At the time of catalog publication, the Stanford TNG server was capable of chromosome localization only on chromosomes 2, 4, 7 and 21. Chromosome localization of markers from the GeneBridge4 panel may be performed by accessing the server at <http://www-genome.wi.mit.edu>. RH mapping involves the statistical analysis of several to many markers to determine the relative order of the markers with respect to one another. RH mapping can be achieved using statistical programs that will provide the best map along with a measure of the relative likelihood of one order versus another.

This type of analysis has been shown to successfully generate the order of markers on the RH map that is significantly more likely than any alternative order. Two statistical programs for RH mapping can be downloaded from the World Wide Web free of charge. SAMapper was produced at the Stanford Human Genome Center and be downloaded at <http://www-shgc.stanford.edu/Mapping/SAMapper/index.html> RHMAP was written by Michael Boehnke at the University of Michigan and can be downloaded at <http://www.sph.umich.edu/group/statgen/software>. A comprehensive web page regarding radiation hybrid mapping, with links to web sites with analysis software and other information, can be found at <http://linkage.rockefeller.edu/tara/rhmap/>

Transfection protocol for cells

N\$ neuroblastoma lines were seeded in Lab Tek chambered coverglass (Nalgene Nunc International) and transfected with pEGFP (control), pGI3303 and pGI3305 using lipofectamine (Life Technologies BRL). After 24-48 hours, the chambered coverglasses were placed on an inverted fluorescence microscope where GFP fluorescence could be visualized in living cells. The

- 61 -

details of this method have been described in
PCT/EP96/02311.

5 **Microscopy and fluorescence staining using
 phalloidin**

 have been described earlier (EP97/06956).

SEQUENCE LISTING

10

Seq ID No 1 is a nucleic acid sequence of Hs unc-53/1
and lacking the nucleotides from position 2873 to 3043
shown in Fig. 1a.

15

Seq ID No. 2 is a nucleic acid sequence of Hs unc-53/1
and lacking the nucleotides from position 3098 to 3121
shown in Figure 1a.

20

Seq ID no. 3 is a nucleic acid sequence of Hs-unc-53/1
and lacking the nucleotides from position 3518 to 3526
of the sequence identified in Fig. 1a.

25

Seq ID No. 4 is an amino acid sequence of Hs-unc-53/1
protein and lacking the amino acids from position 958
to 1014 of the sequence identified in Fig. 1b

30

Seq ID No. 5 is a amino acid sequence of Hs-unc-53/1
protein and lacking the amino acids from position 1033
to 1040 of the sequence identified in Fig. 1b.

35

Seq ID No. 6 is a amino acid sequence of Hs-unc-53/1
protein and lacking the amino acids from position 1173
to 1175 of the sequence identified in Fig. 1b.

Seq ID No. 7 is a nucleotide sequence encoding Hs-
unc-53/2 and lacking the nucleotides from position
5425 to 5433 of the sequence illustrated in Fig. 1c.

Seq ID No. 8 is a nucleotide sequence encoding Hs-unc-53/2 and lacking the nucleotides from position 5924 to 6024 of the sequence illustrated in Fig. 1c.

5 Seq ID No. 9 is a nucleotide sequence encoding Hs-unc-53/2 and having the sequence of variant 1 illustrated in Fig. 1c.

10 Seq ID No. 10 is a nucleotide sequence encoding Hs-unc-53/2 and having the sequence of variant 2 illustrated in Fig. 1c.

15 Seq ID No. 11 is a nucleotide sequence encoding Hs-unc-53/2 and having the sequence of variant 3 illustrated in Fig. 1c.

20 Seq ID No. 12 is a nucleotide sequence encoding Hs-unc-53/2 and having the sequence of variant 1 illustrated in Fig. 1c. and lacking the nucleotides from position 5425 to 5433 of the sequence illustrated in Fig. 1c.

25 Seq ID No. 13 is a nucleotide sequence encoding Hs-unc-53/2 and having the sequence of variant 1 illustrated in Fig. 1c. and lacking the nucleotides from position 5924 to 6024 of the sequence illustrated in Fig. 1c.

30 Seq ID No. 14 is a nucleotide sequence encoding Hs-unc-53/2 and having the sequence of variant 2 illustrated in Fig. 1c. and lacking the nucleotides from position 5425 to 5433 of the sequence illustrated in Fig. 1c.

35 Seq ID No. 15 is a nucleotide sequence encoding Hs-unc-53/2 and having the sequence of variant 2 illustrated in Fig. 1c. and lacking the nucleotides

from position 5924 to 6024 of the sequence illustrated in Fig. 1c.

- 5 Seq ID No. 16 is a nucleotide sequence encoding Hs-unc-53/2 and having the sequence of variant 3 illustrated in Fig. 1c. and lacking the nucleotides from position 5425 to 5433 of the sequence illustrated in Fig. 1c.
- 10 Seq ID No. 17 is a nucleotide sequence encoding Hs-unc-53/2 and having the sequence of variant 3 illustrated in Fig. 1c. and lacking the nucleotides from position 5924 to 6024 of the sequence illustrated in Fig. 1c.
- 15 Seq ID No. 18 is an amino acid sequence of Hs-unc-53/2 protein and lacking the amino acids from position 1776 to 1778 of the sequence identified in Fig. 1d
- 20 Seq Id No. 19 is an amino acid sequence of variant 1 of Hs-unc-53/2 sequence illustrated in Fig. 1d.
- Seq Id No. 20 is an amino acid sequence of variant 2 of Hs-unc-53/2 sequence illustrated in Fig. 1d.
- 25 Seq Id No. 21 is an amino acid sequence of variant 3 of Hs-unc-53/2 sequence illustrated in Fig. 1d.
- Seq Id No. 22 is an amino acid sequence of variant 1 of Hs-unc-53/2 sequence illustrated in Fig. 1d and lacking the amino acids from position 1776 to 1778 of the sequence identified in Fig. 1d.
- 30 Seq Id No. 23 is an amino acid sequence of variant 2 of Hs-unc-53/2 sequence illustrated in Fig. 1d and lacking the amino acids from position 1776 to 1778 of the sequence identified in Fig. 1d.
- 35

- 64 -

Seq Id No. 24 is an amino acid sequence of variant 3 of Hs-unc-53/2 sequence illustrated in Fig. 1d and lacking the amino acids from position 1776 to 1778 of the sequence identified in Fig. 1d.

5

Seq ID No. 25 is a nucleotide sequence encoding Hs-unc-53/3 as illustrated in Figure 1e.

10 Seq ID No. 26 is a nucleotide sequence encoding Hs-unc-53/3 as illustrated in Figure 1e and lacking the nucleotides from position 3795 to 4283 of the sequence identified therein.

15 Seq ID No. 27 is a nucleotide sequence encoding Hs-unc-53/3 as illustrated in Figure 1e and lacking the nucleotides from position 4284 to 4325 of the sequence identified therein.

20 Seq ID No. 28 is a nucleotide sequence encoding Hs-unc-53/3 as illustrated in Figure 1e and lacking the nucleotides from position 3795 to 4325 of the sequence identified therein.

25 Seq ID No. 29 is a nucleotide sequence encoding Hs-unc-53/3 as illustrated in Figure 1e and lacking the nucleotides from position 5153 to 5173 of the sequence identified.

30 Seq ID No. 30 is a nucleotide sequence encoding Hs-unc-53/3 as illustrated in Figure 1e and lacking the nucleotides from position 5343 to 5408 of the sequence identified.

35 Seq ID No. 31 is a nucleotide sequence encoding Hs-unc-53/3 having the sequence of variant 1 illustrated in Fig. 1e.

- 65 -

Seq ID No. 32 is a nucleotide sequence encoding Hs-unc-53/3 having the sequence of variant 1 illustrated in Fig. 1e and lacking the nucleotides from position 3795 to 4283 of the sequence identified therein.

5

Seq ID No. 33 is a nucleotide sequence encoding Hs-unc-53/3 having the sequence of variant 1 illustrated in Fig. 1e and lacking the nucleotides from position 4284 to 4325 of the sequence identified therein.

10

Seq ID No. 34 is a nucleotide sequence encoding Hs-unc-53/3 having the sequence of variant 1 illustrated in Fig. 1e and lacking the nucleotides from position 3795 to 4325 of the sequence identified therein.

15

Seq ID No. 35 is a nucleotide sequence encoding Hs-unc-53/3 having the sequence of variant 1 illustrated in Fig. 1e and lacking the nucleotides from position 5153 to 5173 of the sequence identified therein.

20

Seq ID No. 36 is a nucleotide sequence encoding Hs-unc-53/3 having the sequence of variant 1 illustrated in Fig. 1e and lacking the nucleotides from position 5343 to 5408 of the sequence identified therein.

25

Seq ID No. 37 is an amino acid sequence of Hs-unc-53/3 protein as identified in the sequence of Fig. 1f.

30

Seq ID No. 38 is an amino acid sequence of Hs-unc-53/3 protein as identified in the sequence of Fig. 1f and lacking the amino acid residues from position 1326 to 1413 of the sequence identified therein.

35

Seq ID No. 39 is an amino acid sequence of Hs-unc-53/3 protein as identified in the sequence of Fig. 1f and lacking the amino acid residues from position 1414

- 66 -

to 1427 of the sequence identified therein.

Seq ID No. 40 is an amino acid sequence of Hs-unc-53/3 protein as identified in the sequence of Fig. 1f and lacking the amino acid residues from position 1703 to 1709 of the sequence identified therein.

Seq ID No. 41 is an amino acid sequence of Hs-unc-53/3 protein as identified in the sequence of Fig. 1f and lacking the amino acid residues from position 1768 to 1788 of the sequence identified therein.

Seq ID No. 42 is an amino acid sequence of Hs-unc-53 of variant 1 identified in Figure 1f.

Seq ID No. 43 is an amino acid sequence of Hs-unc-53 of variant 1 identified in Figure 1f and lacking the amino acid residues from position 1326 to 1413 of the sequence identified therein.

Seq ID No. 44 is an amino acid sequence of Hs-unc-53 of variant 1 identified in Figure 1f and lacking the amino acid residues from position 1414 to 1427 of the sequence identified therein.

Seq ID No. 45 is an amino acid sequence of Hs-unc-53 of variant 1 identified in Figure 1f and lacking the amino acid residues from position 1703 to 1709 of the sequence identified therein.

Seq ID No. 46 is an amino acid sequence of Hs-unc-53 of variant 1 identified in Figure 1f and lacking the amino acid residues from position 1768 to 1788 of the sequence identified therein.

- 67 -

CLAIMS

1. A vertebrate protein homologue of a UNC-53 protein of C. elegans, which protein comprises an amino acid sequence having one or more of sequence blocks A, B, C, D, E, F, G, or H as illustrated in figure 4 or which differs from said blocks in conservative amino acid changes.
2. A vertebrate protein homologue of UNC-53 protein of C. elegans or a functional equivalent, derivative or bioprecursor therefor having an amino acid sequence encoded by the nucleotide sequence illustrated in figure 1(e) or the sequence of Figure 1 e having nucleotide region from position 1 to 288 replaced with the sequence of variant 1 illustrated in Figure 1e and or which sequences further lack any of the sequences form 3795 to 4283, 4284 to 4325, 5153 to 5173 or 5343 to 5408.
3. A vertebrate protein homologue of UNC-53 protein of C. elegans having an amino acid sequence as illustrated in figure 1(f) or an amino acid sequence which differs from said amino acid sequence illustrated in figure 1(f) by the replacement of amino acids 1 to 81 with the sequence of variant 1 in figure 1f and /or including deletions from position 1326 to 1413, 1414 to 1427, 1703 to 1709 or 1768 to 1788, or which differs from said sequences in one or more conservative amino acid changes.
4. A cDNA molecule encoding a vertebrate homologue of UNC-53 protein of C. elegans according to any of claims 1 to 3.
5. A cDNA molecule according to claim 4 which cDNA comprises the sequence of nucleotides illustrated

- 68 -

in figure 1(e).

6. A nucleic acid molecule capable of hybridising to the cDNA sequences according to claims 4 or 5 under high stringency conditions.

7. A DNA expression vector which comprises a cDNA molecule as claimed in claim 4 or 5.

8. A vector according to claim 7 which comprises a promoter of C. elegans UNC-53 protein or a vertebrate homologue thereof according to any of claims 1 to 7.

9. A vector according to claim 8 wherein said promoter sequence is derived from a gene encoding a mouse or human homologue of a UNC-53 protein of C. elegans.

10. A vector according to any of claims 7 to 9 which further comprises a sequence encoding a reporter molecule.

11. A vector according to claim 10 wherein said reporter molecule is a fluorophore.

12. A host cell transformed or transfected with the vector of any of claims 7 to 11.

13. A host cell transformed or transfected with the vector of claims 10 or 11.

14. A host cell according to claim 12 or 13 which cell comprises a prokaryotic cell, such as a bacterial cell or a eukaryotic cell such as a fungal, and animal, a plant or an insect cell.

- 69 -

15. A transgenic cell, tissue or organism comprising a transgene capable of expressing a protein according to any of claims 1 to 3.

5 16. A transgenic cell, tissue or organism according to claim 15 which comprises any of a COS cell, Hep G2, MCF-7 cell, N4 mouse neuroblastoma cell, a NIH3Tf cell, or colorectal carcinoma or human derived cells.

10 17. A transgenic cell, tissue or organism according to claim 15 or 16 wherein said transgene comprises a vector according to any of claims 7 to 11.

15 18. A transgenic cell, tissue or organism according to claim 15 or 17 wherein said transgene comprises a vector according to claim 10 or 11.

20 19. A transgenic cell, tissue or organism according to any of claims 15 to 17 wherein said organism comprises any of an insect, a fungus, a non-human mammal, a plant or a nematode worm.

25 20. A method of producing a mutant vertebrate non-human organism which mutation affects cell behaviour or the regulation of cell motility or the shape or the direction of cell migration, which method comprises inducing a mutation in the wild type gene encoding the vertebrate homologue of an UNC-53
30 C. elegans protein.

 21. A vertebrate protein homologue of an UNC-53 protein of C. elegans, according to any of claims 1 to 3 for use as a medicament.

35

 22. Use of a vertebrate protein homologue of an UNC-53 protein of C. elegans, according to any of

- 70 -

claims 1 to 3 in the manufacture of a medicament for promoting neuronal regeneration, revascularisation, wound healing or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease or autoimmune diseases such as rheumatoid arthritis and sclerosis.

23. A pharmaceutical composition comprising a vertebrate homologue of an UNC-53 protein of C. elegans, according to any of claims 1 to 3 together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

24. A nucleic acid or cDNA molecule according to any of claims 4 to 6 or a functional fragment thereof for use as a medicament.

25. Use of nucleic acid or cDNA molecule according to any of claims 4 to 6 in the manufacture of a medicament to promote neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease or autoimmune diseases such as rheumatoid arthritis and sclerosis.

26. A pharmaceutical composition comprising a nucleic acid or cDNA molecule according to any of claims 4 to 6 and a pharmaceutically acceptable carrier, diluent or excipient therefor.

27. A method of determining whether a compound is an inhibitor or enhancer of the regulation of cell behaviour, growth, cell shape or motility or the direction of cell migration, which method comprises contacting said compound with a host cell according to claim 12 or 14 or a transgenic cell as claimed in any of claims 15 to 18 and screening for a phenotypic

- 71 -

change in said cell.

28. A method according to claim 27 wherein said phenotypic change to be screened is a change in cell growth, or shape or a change in cell motility or filopodia outgrowth, ruffling behaviour, cell adhesion, contact inhibition or the length of neurite growth.

29. A method as claimed in claim 27 wherein said transgenic cell is an N4 neuroblastoma cell and the phenotypic change to be screened is the length of neurite growth.

30. A method as claimed in claim 27 wherein said transgenic cell is an MCF-7 breast carcinoma cell or an NIH3T3 cell and the phenotypic change to be screened is the extent of phagocytosis or contact inhibition.

31. A method of determining whether a compound is an inhibitor or an enhancer of the regulation of cell shape, cell growth or motility or of the direction of cell migration, which method comprises administering said compound to a transgenic organism according to any of claims 15 to 19 or a mutant organism produced according to the method of claim 20 and screening for a phenotypic change in said organism.

32. A compound which is identifiable by the method according to claim 27 as an enhancer of the regulation of cell shape, or growth or motility or the direction of cell migration for use as a medicament.

33. Use of a compound which is identifiable by the method according to claim 27 as an enhancer of the

- 72 -

regulation of cell shape, or growth or motility or the direction of cell migration in the preparation of medicament for promoting neuronal regeneration, revascularisation or wound healing or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease autoimmune diseases such as rheumatoid arthritis or sclerosis.

34. A pharmaceutical composition comprising a compound identified according to the method of any of claims 27 to 31 and a pharmaceutically acceptable carrier, diluent or excipient therefor.

35. A compound which is identifiable by the method according to any one of claims 17 to 31 as an inhibitor of the regulation of cell motility, growth, or shape, or the direction of cell migration, for use as a medicament.

36. Use of a compound according to claim 35 in the manufacture of a medicament for alleviating the spread of disease inducing cells or metastasis or loss of contact inhibition.

37. A pharmaceutical composition comprising the compound as claimed in claim 35, and a pharmaceutically acceptable carrier diluent or excipient therefor.

38. A method of determining whether a compound is an inhibitor or an enhancer of transcription of a gene encoding a vertebrate homologue of UNC-53 protein of C. elegans, according to any of claims 1 to 3 which method comprises the steps of (a) contacting said compound with a cell according to claim 13 or 18 and (b) monitoring the level of said reporter molecule and comparing the results obtained from said monitoring

- 73 -

step with a control comprising a cell according to claims 13 or 18, which cell has not been contacted with said compound.

5 39. A method as claimed in claim 38 wherein said reporter molecule detected is mRNA or green fluorescent protein.

10 40. A compound which is identifiable by the method according to claims 38 or 39, as an enhancer of transcription of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans according to any of claims 1 to 3 or a functional fragment of said gene, for use as a medicament.

15 41. Use of a compound which is identifiable by the method of claims 38 or 39, as an enhancer of transcription of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans according to any of claims 1 to 3 or a functional fragment of said gene, in the manufacture of a medicament for promoting neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neuro-degenerative diseases or acute traumatic injuries or
20 fibriotic disease or autoimmune diseases such as
25 rheumatoid arthritis or sclerosis.

 42. A pharmaceutical composition which comprises the compound of claim 40 and a pharmaceutically
30 acceptable carrier, diluent or excipient therefor.

 43. A compound which is identifiable by the method of claims 38 or 29 as an inhibitor of transcription of a gene coding for vertebrate
35 homologue of a UNC-53 protein of C. elegans according to any of claims 1 to 3 or a functional fragment of said gene for use as a medicament.

- 74 -

44. Use of a compound which is identifiable by the method of claims 38 or 39 as an inhibitor of transcription of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans or a functional fragment of said gene, in the manufacture of a medicament for alleviating spread of disease inducing cells or metastasis or loss of contact inhibition.

45. A pharmaceutical composition which comprises the compound of claim 43 and a pharmaceutically acceptable carrier, diluent or excipient therefor.

46. A kit for determining whether a compound is an enhancer or an inhibitor of the regulation of cell motility, growth or shape or the direction of cell migration which kit comprises at least one transgenic cell as claimed in any one of claims 13 to 17 to be contacted with said compound and at least one cell according to claims 18 to 19 to be used as a control and means for contacting said compound with one of said at least one transgenic cells.

47. A kit for determining whether a compound is an inhibitor or an enhancer of transcription of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans or a functional fragment of said gene which kit comprises at least one cell as claimed in any one of claims 12 to 19 and means for contacting said compound with said cells.

48. A kit for determining whether a compound is an enhancer or an inhibitor of the activity of a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, derivative, fragment or bioprecursor of said vertebrate homologue protein, which kit comprises at least, one vertebrate

- 75 -

mutant non-human organism produced according to the method as claimed in claim 20 or a transgenic organism as claimed in claims 15 to 19 and a wild type of said vertebrate mutant organism.

5

49. A method identifying vertebrate homologues of an unc-53 gene of C. elegans or a functional fragment thereof, which method comprises hybridizing to a DNA library a suitable
10 oligonucleotide sequence of between 15 to 50 nucleotides of the nucleic acid sequence encoding UNC-53 or a functional equivalent, derivative or bioprecursor thereof, under appropriate conditions of stringency to identify genes having statistically
15 significant homology with the cDNA according to any of claims 4 or 5.

50. A method of identifying a protein which is active in the signal transduction pathway of a cell of
20 which a vertebrate homologue of an UNC-53 protein of C. elegans according to any of claims 1 to 3 is a component, which method comprises:

- 25 (a) contacting an extract of said cell with an antibody to the vertebrate homologue of the UNC-53 protein of C. elegans,
- (b) identifying the antibody/vertebrate homologue complex, and
- (c) analysing the complex to identify any protein bound to the vertebrate homologue of
30 UNC-53 protein of C. elegans other than the antibody.

51. A method of identifying a further protein which is active in the signal transduction pathway of
35 a cell of which a vertebrate homologue of an UNC-53 protein according to any of claims 1 to 3 is a component, which method comprises:

- 76 -

(a) forming an antibody to the first identified protein bound to the vertebrate homologue of UNC-53 protein of C. elegans in claim 50,

5 (b) contacting a cell extract with said antibody and identifying the antibody/protein complex,

(c) analysing the complex to identify any further protein bound to the first protein
10 other than the antibody, and

(d) optionally repeating steps (a) to (c) to identify further proteins in said pathway.

15 52. A method of identifying a protein which is active in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein of C. elegans according to any of claims 1 to 3 is a component, which method comprises:

20 (a) contacting an extract of said cell with said vertebrate homologue of an UNC-53 protein of C. elegans,

(b) identifying any vertebrate homologue of UNC-53 protein/protein complex formed and

25 (c) analysing the complex to identify any protein bound to the vertebrate homologue of UNC-53 protein other than the same vertebrate homologue of UNC-53 protein.

30 53. A method according to claim 52 which further comprises contacting a cell extract with any protein identified from step (c) not being the same as the vertebrate homologue of UNC-53 protein used and repeating steps (b) and (c) so as to identify any
35 further protein involved in the signal transduction pathway of said cell.

- 77 -

54. A method of identifying a protein involved in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein of C. elegans is a component which method comprises:

- 5 (a) providing an appropriate host cell having a DNA construct comprising a reporter gene under the control of a promoter regulated by a transcription factor having a DNA binding domain and an activating domain,
- 10 (b) expressing in said host cell a first hybrid DNA sequence encoding a first fusion of a fragment or all of a DNA sequence according to claims 4 or 5 and either said DNA binding domain or the activating domain
- 15 of the transcription factor,
- (c) expressing in the host cell at least one second hybrid DNA sequence encoding a putative binding protein to be investigated together with the DNA binding or activating
- 20 domain of the transcription factor which is not incorporated in the first fusion,
- (d) detecting any binding of the protein being investigated with a protein according to any of claims 1 to 3 by detecting for the
- 25 production of any reporter gene product in said host.

55. A protein identified by the method of any one of claims 50 to 54 for use as a medicament.

30

56. Use of a protein identified by the methods of any one of claims 50 to 54 in the manufacture of a medicament for promoting neuronal regeneration, revascularisation or wound healing, or for treatment

35 of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease or autoimmune diseases such as rheumatoid arthritis and sclerosis.

- 78 -

57. A pharmaceutical composition comprising a protein identified by the methods of any one of claims 50 to 54 and a pharmaceutically acceptable carrier, diluent, or excipient therefor.

5

58. A process for producing a vertebrate homologue of an UNC-53 protein of C. elegans according to any of claims 1 to 3 which process comprises culturing the cells of any of claims 12 to 14 and
10 recovering said vertebrate homologue of UNC-53 protein expressed.

59. A process for producing a vertebrate homologue of an UNC-53 protein of C. elegans according
15 to any of claims 1 to 3 which process comprises culturing an insect cell transfected with a recombinant Baculovirus vector, said vector comprising a DNA insert encoding said vertebrate homologue of UNC-53 protein downstream of the Baculovirus
20 polyhedrin promoter, and recovering the expressed vertebrate homologue of UNC-53 protein.

60. A method of detecting whether a compound is an inhibitor or an enhancer of expression of a
25 vertebrate homologue of an UNC-53 of C. elegans according to any of claims 1 to 3 which method comprises contacting a cell expressing said homologue with said compound and monitoring for a phenotypic change compared to a control cell which has not been
30 contacted with said compound.

61. A method according to claim 60 wherein said cell comprises a cell according to any of claims 12 to 19.

35

62. A method according to claim 60 wherein said cell has undergone loss of contact inhibition.

- 79 -

63. A method according to any of claims 60 to 62 in which the compound to be tested comprises a nucleic acid.

5 64. A method according to claim 63 wherein said nucleic acid sequence comprises an antisense DNA or RNA sequence.

10 65. A method according to claim 64 wherein said mRNA sequence comprises 3' untranslated regions of mRNA encoding for said vertebrate homologue.

15 66. A method according to any of claims 60 to 62 wherein said compound to be tested comprises a protein having an amino acid sequence potentially suitable for inhibiting function of said vertebrate homologue.

20 67. A method according to claim 66 wherein said protein comprises a protein identified according to any of the methods of claims 50 to 54.

25 68. A pharmaceutical composition comprising a compound identified according to any of claims 60 to 67 together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

30 69. A nucleic acid sequence identified according to the method of any of claims 63 to 65 for use as a medicament.

35 70. Use of a nucleotide sequence identified according to the method of any one of claims 63 to 65 in the preparation of a medicament for the treatment of loss of contact inhibition or cancer which is mediated by a vertebrate homologue of an UNC-53 protein of C. elegans.

- 80 -

71. Use of a nucleic acid according to claim 69 in the preparation of a medicament for inhibiting expression of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans.

5

72. An assay for detecting expression of a vertebrate homologue of UNC-53 protein of C. elegans according to any of claims 1 to 3 in a vertebrate cell which assay comprises contacting a cell or an extract thereof with an antibody to said vertebrate homologue, which antibody is linked to a reporter molecule, removing any unbound antibody and monitoring for the presence of said reporter molecule.

10

73. An assay according to claim 72 wherein said reporter molecule is an antibody conjugated with a suitable fluorophore or detectable enzyme.

15
20

74. A method for detecting for expression of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans according to any of claims 1 to 3 which method comprises contacting a probe specific for a nucleic acid or protein sequence coding for or corresponding to said vertebrate homologue according to any of claims 1 to 3 with a cell extract which probe is linked to a reporter and analysing for the presence of said reporter.

25

30

75. A method according to claim 74 wherein said probe comprises a complementary sequence to a region of mRNA transcribed from said gene encoding said vertebrate homologue of UNC-53 protein.

35

76. A method according to claim 75 wherein said complimentary sequence is a 3' or 5' untranslated region of said mRNA.

- 81 -

77. A method according to claims 74 or 76 wherein said reporter comprises a radiolabel.

5 78. A method according to claim 74 wherein said probe comprises an antibody specific for said vertebrate homologue of said UNC-53 protein according to any of claims 1 to 3.

10 79. A method according to claim 78 wherein said reporter comprises an antibody conjugated with a detectable fluorophore or enzyme.

15 80. A method of determining whether a compound is an inhibitor or an enhancer of association of a vertebrate homologue according to any of claims 1 to 3 to microtubules or plus end regions thereof, which method comprises:-

20 (a) contacting said compound with a transgenic cell, tissue or organism expressing UNC-53 protein or said vertebrate homologue and which protein is operably linked to a reporter molecule,
(b) screening for the localisation of said reporter molecule as compared to a cell
25 according to step (a) which has not been contacted with said compound.

30 81. A compound identifiable by the method according to claim 80.

82. A compound according to claim 81 for use as a medicament.

35 83. Use of a compound according to claim 81 as an enhancer of association of said vertebrate homologue with microtubules or the plus end region thereof, for use in promoting neuronal regeneration,

- 82 -

revascularisation or wound healing, or for treating chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease or autoimmune diseases such as rheumatoid arthritis or sclerosis.

5

84. A pharmaceutical composition comprising the compound according to claims 81 or 82 and a pharmaceutically acceptable carrier, diluent or excipient therefor.

10

85. A kit for determining whether a compound is an inhibitor or an enhancer of association of a vertebrate homologue according to any of claims 1 to 3 with microtubules or the plus end regions thereof, which kit comprises at least one transgenic cell expressing said homologue and a reporter molecule or a cell according to any of claims 12 to 19 and at least one cell of the same cell type for use as a control and means for contacting said compound with one of said at least one transgenic cells.

20

86. A composition comprising a vertebrate homologue according to any of claims 1 to 3 linked to a compound identified as an inhibitor or enhancer or association of said vertebrate homologue with microtubules or their plus end regions for use in targeting said compound to said microtubule or the plus end region thereof.

25

87. A composition according to claim 86 which further comprises a cell transformation or transfecting agent.

30

88. A method of targeting a protein to a cell microtubule or the plus end region thereof, which method comprises introducing into a host cell, tissue or organism a transgene comprising a sequence capable

35

- 83 -

of expressing a vertebrate homologue according to any of claims 1 to 3, which sequence is operably linked to a sequence encoding said protein to be targeted such that a chimeric protein is expressed and which results
5 in targeting said protein to said microtubule or a plus end region thereof.

89. A method of identifying a molecule which covalently modifies a vertebrate homologue of UNC-53
10 according to any of claims 1 to 3 which method comprises:

- a) contacting an extract from a cell expressing said vertebrate homologue with a mixture of enzymes comprising candidate modifying enzymes in
15 the presence of an inhibitor or covalent modification of a protein,
- b) identifying any covalently modified UNC-53 protein from step a),
- c) identifying said molecule involved in said
20 modification step.

90. A method according to claim 89, wherein said indicator comprises ^{32}p .

25 91. A method of identifying a compound which alleviates or enhances the toxicity of a vertebrate homologue according to any of claims 1 to 3, which method comprises contacting said compound with a cell, tissue or organism according to claim 18, and
30 monitoring for the presence of said reporter molecule adjacent said microtubules or the plus end regions thereof.

92. A vertebrate homologue of UNC-53 protein of
35 C.elegans or a functional equivalent, derivative or bioprecursor therefor encoded by the nucleotide sequence in Figure 1a and which nucleotide sequence is

- 84 -

lacking in any of the nucleotide regions from position 2873 to 3043, 3098 to 3121 or 3518 to 3526.

93. A vertebrate homologue of UNC-53 protein of
5 C.elegans or a functional equivalent, derivative or
bioprecursor therefor having an amino acid sequence as
illustrated in Figure 1b and lacking in one or more of
the regions from residues 958 to 1014, 1033 to 1040 or
1173 to 1175, or which differs from said amino acid
10 sequences in one or more conservative amino acid
changes.

94. A vertebrate homologue of UNC-53 protein of
C.elegans or a functional equivalent, derivative or
15 bioprecursor therefor encoded by the nucleotide
sequence in Figure 1c and which nucleotide sequence
has from sequence position 1 to 366 replaced with any
of the sequences identified as variants 1 to 3 of
Figure 1c and/or which sequences lack the region from
20 position 5624 to 6024.

95. A vertebrate homologue of UNC-53 protein of
C.elegans or a functional equivalent, derivative or
bioprecursor therefor having an amino acid sequence
25 identified in Figure 1d or the sequences of any of
variants 1 to 3 replacing the amino acids from
position 1 to 89 of the sequence of Figure 1d and/or
which sequence is lacking the amino acid sequence from
position 1776 to 1778.

30

96. Plasmid pG313303 deposited under accession
number LMBP 3936.

97. Plasmid pG13305 deposited under accession
35 number LMBP 3937.

1156

Figure 1a. Nucleotide sequence of Hs-unc-53/1

CATGCTGCCAAGCGCGCCAAGGCGCCCGCGGCGGCGGCGGCATGGCCAAGGCCAGCGCGGCTGAGCTGAAGGT 75
 CTTCAAGTCCCGGCAGCGTGGACAGCCGTGTCCCGGCGGGCGCGCCGCCCTCCAACCTGCGCAAGCAGAAGTCACT 150
 CACCAACCTCTCTTTTCTCACGGACTCCGAGAAAAAGCTGCAGCTTTATGAGCCCGAATGGAGCGACGATATGGC 225
 CAAGGCGCCCAAGGCTTAGGCAAGGTGGGGTCCAAGGGCCGTGAAGCTCCGCTGATGTCCAAGACGCTGTCCAA 300
 GTCGGAGCACTCGCTCTTCCAGGCCAAGGGCAGCCCGCGGGCGGTGCCAAGACCCCCCTGGCTCCGCTCGCGCC 375
 CAACCTGGGAAAGCCGAGCCGGATCCCTCGAGGACCCTATGCGGAGGTCAAGCCGCTCAGCAAGGCGCCTGAAGC 450
 GGCCGTGAGCGAAGATGGCAAATCGGACGACGAGCTGCTCTCCAGCAAGGCCAAGGCGCAAAAGAGCTCTGGGCC 525
 TGTCCCTCTGCCAAGGGCCAGGAGGAGCGCGCCTTCTCAAGGTGGACCCGAGCTGGTGGTGACCGTGCTGGG 600
 AGACCTGGAGCAGCTGCTCTTCAGCCAGATGCTGGACCCAGAGTCCCAGAGAAAGAGGACAGTGCAGAATGTCTT 675
 GGATCTCCCGCAGAACCTGGAAGAGACCATGTCCAGCCTGCGAGGGTCCCAGGTGACTCACAGCTCCCTGGAGAT 750
 GACCTGCTACGACAGCGATGATGCCAACCCACGACGCTGTCCAGCCTCTCCAACCGCTCGTCCCTCTGTTCATG 825
 GCGCTATGGCCAGTCCAGTCCGCGGCTGCAGGCTGGTGACGCGCCCTCTGTGGGTGGGAGCTGCCGCTCGGAGGG 900
 GACGCCCCCTGGTACATGCACGGCGAACGGGCCACTACTCCACACCATGCCCATGCGCAGCCCCAGCAAGCT 975
 CAGCCATATCTCCCGCCTGGAGCTGGTTCGAATCCCTGGACTCGGATGAGTGGACCTCAAGTCCGGCTACATGAG 1050
 CGACAGTACCTCATGGGCAAGACCATGACGGAGGATGATGACATCACTACCGCTGGGATGAAAGCAGCTCCAT 1125
 CAGTAGTGGACTCAGCGATGCCTCAGACAATCTCAGTTCAGAAGAATTCAATGCCAGCTCCTCACTCAATCCCT 1200
 CCCAAGTACTCCCACTGCTTCTCGCAGGAACCTCAACAATAGTGCTACGCACAGACTCAGAGAAGCGCTCACTGGC 1275
 AGAAAGTGGGCTGAGCTGGTTTAGTGAATCAGAGGAGAAAGCCCCATAAAAACTGGAGTACGACAGTGGTAGCCT 1350
 GAAGATGGAACCTGGGACTTCTAAGTGGCGGAGGGAGCGGCTGAGAGCTGTGATGATTCAAGGTTGAGAGA 1425
 ACTGAAAAAGCCCATCAGCCTGGGCCACCTTGGTTCCCTGAAGAAGGGCAAGACCCACCTGTGGCTGTAACCTC 1500
 CCCCATCACTCACACAGCCAGAGTGCCTCAAAGTTCGAGGCAAACTGAGGGCAAAAGCTACAGACAAGGGTAA 1575
 GCTTGCAGTGAAGAATACTGGGCTCCAACGCTCCTCTGTATGCTGGTTCGGGACCGCTGAGTGTGCTAAGAA 1650
 GCCCCCCCTCGGCATTGCTCGCCCCCTCCACTTCGGGATCCTTTGGGTACAAGAAGCCTCCTGCTGCCACAGGCAC 1725
 AGCCACTGTTCATGCAAACTGGTGGTTTCAGCCACTCTCAGCAAGATCCAGAAGTCTCAGGCATCCCTGTCAAGCC 1800
 AGTAAATGGGCGCAAGACTAGCTTAGATGTTTCCAACAGTGCAGAGCCAGGATTCTTGGCTCCTGGAGCCGTTT 1875
 TAACATCCAGTACCGCAGCCTGCCCCGCGCCAGCCAAGTCAAGTTCTATGAGCGTGACCGGCGGGGCGGGTGGAC 1950
 TCGCCCTGTGAGCAGCAGCATTGACCCAGTCTCCTCAGCACCAGCAGGAGGCCTTACGCCTTCAGACTGAA 2025
 GGAGCCTACCAAGGTAGCCAGTGGGCGGACCACTCCAGCCCTGTCAATCAGACAGATCGGGAAAAGGAGAGGC 2100
 CAAAGCCAAAGCAGTGGCCTTGGACTCAGACAACATCTCCTTGAAGAGTATTGGCTCCCCAGAAAGTACTCCCAA 2175
 GAACCAAGCAAGCCACCCACAGCCACCAAGTGGCAGAGCTGCCACCAACCCCTCTCAGGGCCACAGCGAAGAG 2250
 CTTTGTCAAACACCCCTCACTAGCCAATCTTGACAAGGTCAACTCCAACAGTCTGGATCTACCATCATCCAGTGA 2325
 TACCACCATGTCTCAAAGGTCCAGATCTGCATGCTACAAGCTCAGCATCTGGGGGCCCTCTCCCTTCCCTGCTT 2400
 CACCCCTCCCAAGAAAGGCTCAGGTACCACTTCAATATTAAGTCTCAGCCAGCTTCTCCAGGGCCTGGAGCTAATGAGTGGTTT 2475
 CAGTGTGCCAAAAGAGACCCGATGTACCCCAAACCTCTCAGGCCTGCACAGGAGCATGGAGTCCCTCCAGATGCC 2550
 AATGAGCCTCCCCAGTGCCTTCCCCAGCAGTACTCCCGTCCCCACCCACCTGCTCCCCCTGCTGCTCCACAGA 2625
 AGAAGAGACGGAAGAGCTGACTTGGAGTGGAGGCCAGAGCTGGGCAACTGGACAGTAATCAGCGGATCGGAA 2700
 CACTCTTCCCAAGAAAGGCTCAGGTACCACTTCAAGTCCAGGAGGAGACCAAGGAGAGGCGACATTCCCATAC 2775
 CATTGGTGGGCTGCCTGAATCCGATGACCAGTCAAGAGCTGCCTTCTCCCCCTGCACTTCCCATGTCTCTGAGTGC 2850
 AAAGGGCCAACTTACCAACATAgtagtcccactgcgccaccacgccaagaatcacccgctccaacagcatccc 2925
 caccacagagggcgcccttcgagctgtacagcggtcccaaatggggagcaccctgtccctggccgagagacccaa 3000
 gggaaatgattcggtcaggatcccttcgagacccacggacgatGTTACGGCTCAGTGCTGCTCCCTCCAG 3075
 TGCTCTCCCACTCACTCTCAgctgaggagaggaatctgagCAATCCGGAAGCTTCGTAGGGAAGTGA 3150
 ATCATCCCAGGAAAAAGTGGCCACCTTGACGTCTCAGCTTTCTGCCAATGCTAATCTGGTGGCTGCTTTTGAGCA 3225
 GAGCCTGGTGAATATGACATCCCGCCTGCGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGA 3300
 TTTGCGAGAAACCATAGACTTTCTGAAGAAAAAGAACTCTGAGGCCCCAGGCAGTCACTCAGGAGCCCTTAATGC 3375
 CTCAGAAACCACACCCAAAGAACTTCGGATCAAGAGACAAAAAATCTCAGATAGCATCTCAAGCCTCAACAGCAT 3450
 CACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAGAAAAAAGAGTTGGgtctatga 3525
 gCTTCGAAGTTCTTCAACAAAGCGTTCAAGTATAAAAAAGGGGCCAAAGTCAGCTTCTCATACTCGGATATAGA 3600
 GGAGATTGCTACACCCGACTCTTCAGCCCCCTCATCCCCAAACTACAGCATGGTTCTACAGAGACTGCTTCACC 3675
 CTCCATCAAGTCTCCACCTYGTCTCCGTGGGCACTGATGTACCGAGGGCCCTGCTCACCCAGCCCCCACAC 3750
 TAGGCTGTTTCCATGCAAAATGAGGAGGAGGAGCCAGAGAAGAGGAGGTATCGGAGCTGCGCTCTGAGCTATGGGA 3825
 GAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCACCAACTGGATCAGCTTCGGGAGAC 3900
 CATGCACAACATGCAGTTGGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCTC 3975
 ATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCTGCATTATCTTCCCCACGCCCTCCCTTAGGCTGGCACT 4050
 CACCAATTCTTCCGCCCCAGTCTTGCAGACACAGCACTGTTCACCCATGGATGGCATCAGTACTTGTGGTCCAAA 4125
 GGAGGAAGTGACCTCCGGGTGGTGGTGGAGATGCCCCCGCAGCACATCATCAAAGGGGACTTGAAGCAGCAGGA 4200
 ATTCTTCTGGGCTGTAGCAAGGTCAAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCAAGTGTT 4275
 CAAGGACTATATTTCTAAATGGACCCAGCTCTACCTTGGGACTAAGCACTGAGTCCATCCATGGCTACAGCAT 4350
 CAGCCACGTGAAACAGTGTGGATGACAGAGCCCCAGATGCTTCCCTTGGCGTCGAGGTGTCATAACATATC 4425
 AGTCTCCCTCAAAGGCTGAAGGAGAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGAT 4500
 GCAGCACTACATAAGCCTCCTGCTGAAGCACCGGCCCTCGTCTCTCGGGCCCCAGCGGCACGGGCAAGACCTA 4575
 CCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTACAGAGGGGCATCGTCAGCACCTT 4650
 CAACATGCACCAGCAGTCTTGCAAGGATCTGCAAGTGTATCTTTCCAACCTAGCCAACCATAGTACCGGAAAC 4725
 AGGAATTGGGATGTGCCCTGGTGATTCTATTGGATGACTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAA 4800
 TGGGGCCCTCACCTGCAAGTATCATAAATGTCCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACC 4875

2/56

Figure 1a (CONTINUED)

CAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACCTTCTCCAACAACGTGGAGCCAGCCAATGGCTTCCTGGT 4950
 TCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTCGGGTGCT 5025
 CGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGG 5100
 CCCTTGCTTCTTTCTGTCGTGTCCCATTTGGCATTGAGGACTTCCGGACCTGGTTCATTGACCTGTGGAACAAC TC 5175
 TATCATTCCCTATCTACAGGAAGGAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACCC 5250
 AGTGAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACCTGCCCCC 5325
 ACCCACCCTGGGCCCTCACAGCATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCT 5400
 GGACTCAGATCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCCAATACATTGAGTCTCCAGATCGAGA 5475
 AACCATCCTGGACCCCAACCTTCAGGCAACACTTTAAGGGTTTCGGCAATCACTGTCAACCCCGGACAGCAGAACG 5550
 CTGGCATCAGCTATCTTAGCTCCTCTCCTCTCCCTCTCCTCTTTAGAGCACTGGCTCTCCAGCCCCAGGAGGAGA 5625
 ACAGGAGGGAGGAGGAGATGAAAGAGGAGGGACAGGTTCTTGCTGTACCTTTGAGAACTTCTAGGAAGGAA 5700
 TGGTGGGTGGCGTTTGGGAACCTGTGCCCCCTAAACACATTTACTGGCCTCCTCTAATGACTTTGGGGAAAAGA 5775
 TGATTCTGGGTCTTTCCCTTGACTTCTTTGTTCAATTACAACTCCTGGGCTTTCTGGGGAGGGGTTTCAGAAAAC 5850
 ATCAAAACACTGCAGCAGTTCCTAAATGATTCTCAAGCAACCCTGAGAGAGACAGTCTTGTGAGGGAGATCTG 5925
 GGGGAGGCAGGAAGCTCCTCAGATTTTCTCACAGACCCTTCCCAATTCATCACCCTGCCAACACTCGTCCGGA 6000
 ATTC 6004

In frontal cortex, variants have been found lacking the region from position 2873 to 3043 or the region from residues 3098 to 3121. The region from 3518 to 3526 is absent in cDNA from Hela or colorectal adenocarcinoma tissue. All three regions are indicated in lower case letters in the figure above. Y at position 3696 stands for C or T. Both nucleotides have been found to be present in cDNAs from different origin.

3/56

Figure 1b. Amino Acid sequence of the protein encoded by Hs-unc-53/1 gene. Stretches encoded by the DNA sequences lacking in variants from frontal cortex are in lower case letters (residues 958 to 1014 ; 1033 to 1040 and 1173 to 1175). The x at position 1232 stands for Leucine or Serine, depending on the cDNA of origin.

```

MLPKRAKAPGGGGGMAKASAAELKVFKSGSVDSRVPGGPPASNLRKQKSLTNLSFLTDSKKLQLYEPEWSDDMA 75
KAPKGLGKVGSKGREAPLMSKTLSEHSLFQAKGSPAGGAKTPLAPLAPNLGKPSRIPRGPYAEVKPLSKAPEA 150
AVSEDKSDDELLSSKAKAQKSSGPVPSAKGQEERAFKVDPELVVTVLGDLEQLLFSQMLDPESQQRKRTVQNVL 225
DLRQNLREETMSSLRGSQVTHSSLEMTCYDSDANPRSVSSLSNRSSPLSWRYGQSSPRLQAGDAPSVGGSCRSEG 300
TPAWYMHGERAHYSHTMPMRSPSKLSHISRLELVESLDSDEVLDKSGYMSDSDLMGKTMTEDDDDITGWDDESSI 375
SSGLSDASDNLSSSEFNASSSLNSLPSTPTASRRNSTIVLRDSEKRS LAESGLSWFSESEEKAPKKLEYDSGSL 450
KMPEGTSKWRRRERPESCDSSKGGELKKPISLGHPGSLKKGKTPPVAVTSPITHTAQSALKVAGKPEGKATDKGK 525
LAVKNTGLQRSSSDAGRDRLSDAKKPPSGIARPSTSGSFGYKKPPPATGTATVMQTGGSATLSKIQKSSGIPVKP 600
VNGRKTSLDVSNSAEPGFLAPGARSNIQYRSLPRPAKSSSMSVTGGRGGPRPVSSSIDPSLLSTKQGGTTPSRK 675
EPTKVASGRTPAPVNTDREKEKAKAKAVALDSDNISLKSIGSPESTPKNQASHPTATKLAELPPTPLRATAKS 750
FVKPPSLANLDKVNNSLDLPSSSDTTHASKVPDLHATSSASGGPLPSCFTPSPAPILNINSASFSGLELMMSGF 825
SVPKETRMPYKLSGLHRSMESLQMPMSLPSAFPSSTPVPTPPAPPAAPTEETEELTWGSPRAGQLDSNQDRN 900
TLPKKGLRYQLQSQEETKERRHSHTIGGLPESDDQSELPSPPALPMSLSAKGQLTNIVSPTAATTPTITRSNSIP 975
theaafelysgsqmgstlslaerpkgmirsgsfrdptddVHGSVLSLASSASSTYSsaeermqseQIRKLRELE 1050
SSQEKVATLTSQLSANANLVAAFEQSLVNMTSRLRLAETAEEKDTLELDLRETIDFLKKKNSEAAVIGQALNA 1125
SETTPKELRIKRONSSDSISSLNSITSHSSIGSSKDADAKKKKKSwvyeLRSSFNKAFSIKKGPKSASSYSDIE 1200
EIATPDSSAPSSPKLQHGSTETASPSIKSSTxSSVGTDTVTEGPAHPAPHTRLFHANEEEEPEKKEVSELSELWE 1275
KEMKLTDIRLEALNSAHQLDQLRETMHNMQLVLDLLKAENDRLKVAPGPSSGSTPGQVPGSSALSSPRSLGLAL 1350
THSFGPSLADTDLSPMDGISTCGPKKEEVTLRVVVRMPPQHI IKGDLKQOEFFLGCSKVSQKVDWKMLDEAVFQVF 1425
KDYISKMDPASTLGLSTESIHGYSISHVKRVLDAEPPPEMPPCRGVNNISVSLKGLKEKCVDSLVEFETLIPKPM 1500
QHYISLLKKHRRVLVSGPSGTGKTYLTNRLAEYLVERSGREVTEGIVSTFNMHQQSKDLQLYLSNLANQIDRET 1575
GIGDVPLVILLDDLSEAGSISELVNGALTCKYHKCPYIIIGTTNQPVKMTPNHGLHLSFRMLTFSNNVEPANGFLV 1650
RYLRRKLVESDSDINANKEELLRLVDWVPKLWYHLHTFLEKHSTSDFLIGPCFFLSCPIGIEDFRTWFI DLWNNS 1725
IIPYLQEGAKDGIKVHGQKAAWEDPVEWVRDTLPWPSAQDQSKLYHLPPPTVGPHSIASPPEDRTVKDSTPSSL 1800
DSDPLMAMLLKLQEAANYIESPDRETILDPNLQATL 1835

```

4/56

Figure 1c. Nucleotide sequence of the Hs-unc-53/2 2 gene

TAAGGCCCGGCGCCTGCTCTGCTACCCGCGCTGCCTTTAGCGGTGCGCCCCCGCGCGCTGCCAGGGACGTGCTG 75
 GGAAAGCCCAAGCCCCGGGAGAGATGCCGGCCATCCTGGTGCCTCCAAAATGAAGTCGGGACTGCCCAAACCC 150
 GTGCACAGCGCCGCGCCCATCCTGCACGTGCCCCCGGCGCGGGCGGGCCCCAGCCCTGCTACCTGAAGTTGGGA 225
 AGCAAGGTGGAGGTGAGCAAGACCACCTATCCTAGCCAGATCCCCCTGAAATCCAGGTGCTGCAGGGGCTGCAG 300
 GAGCCAGCGGGGAGGGGCTCCCGCTGCGGAAGAGCGGCTCGGTGGAAAACGGGTTCGATACCCAGATCTACACA 375
 GACTGGGCCAATCATTACCAAAATCCGGCCACAAGCGTCTCATCAAGGATCTCCAGCAAGATGTGACAGAT 450
 GCGCTCCTCTGGCCAGATTATCCAGTTGTGGCAAATGAAAAGATTGAAGACATCAATGGCTGTCCGAAGAAC 525
 AGATCCCAAATGATTGAAAACATAGATGCCCTGCTTGAATTTCTGGCAGCTAAGGGAATAAATATCCAGGGGCTG 600
 TCTGCAGAAGAGATCAGGAATGGAAACCTCAAGGCCATTCTAGGCCTCTCTTCAGCCTCTCCCGATACAAGCAG 675
 CAGCAGCAGCAGCCCCAGAAGCAGCACCTCTCCTCACCTCTGCCGCCCGCGGTATCCAGGTGGCCGGGGCCCC 750
 TCCAGTGCAGGCTGGCCACCCCTCAGCAGCAGGTGCCAGTCACTCCCCAAGCCCCGTGCCAGGCTCACCAGCCA 825
 GCGCCATGCAGCAGTCAAAAGCACAAGCTGAAATGCAGTCCAGACTTTCAGGTCTACCGCGAGGGTATCCGCT 900
 GCAGGCAGCGAGGCCAAAACACGCGGAGGGTCAACTACTGCTAACACCGACGCAGCCAGAGCTTTAACTAT 975
 GATAAATCCAAACAGTCACCTCCCCACCCCCACCGCCAAGCAGCCAGAGAAAGAGCCTTTGGCAAGTTCAGCC 1050
 TCCTCCCACCCCGGAATGAGTGACAATGCACCTGCTTCTTGGAGAGCGGCAGCAGCTCCACCCCTACTAATTGC 1125
 AGTACCTCTCGGCCATCCCGCAGCCCGGTGCAGCCACCAAGCCTTGGCGCAGCAAAATCCCTCAGCGTGAAGCAG 1200
 AGTGCCACGGTATCCATGCTCTCGGTCAAGCCTCCTGGGCGTGAAGCCCCAGGCCACACCTGAAGCCATGAAG 1275
 CCGGCCCCCAACAATCAGAAATCCATGCTGGAAAAGCTGAACTTTTCAACAGTAAAGGGGGCTCAAAGGCAGGT 1350
 GAGGGGCGGGGTCCCGGACACAAGCTGTGAGCGGCTGGAGACTCTGCCAGCTTCGAAGAGAGCGGAGGCTG 1425
 GAGGCGCCATGCTGCTACCCACCGTGGGCTTGTCTCAGCAGCCCCAGCTTCAGCTCAAGGCCATGGCC 1500
 CAGAGGACTTTTATCGGCACTGACCAACAAGAGAGTTCCTCTGAAAAGGCAATGAGAAAGAGAGAGAAACAA 1575
 CAGCGGGAGAAGGATAAGGAGAAAAGCAAGGACCTTGCCAAGAGAGCCTCTGTGACGGAGAGGCTGGACCTCAAG 1650
 GAGGAGCCAAAAGAAGACCCAGTGGAGCAGCTGTGCCGAGATGCCAAAAAAGTCTCCTCAAGATTGCCAGCTTC 1725
 ATCCCCAAAGGGGGGAAGCTCAACAGTGCCAAGAAGGAGCCCCATGGCCCCCTTCCACAGTGGAAATACCAAACCA 1800
 GGAATGAAGAGCATGCCCGGGAATCCCCAAGTGCCCGCAGCCCTTCCAAGGAAGGGGAGCGGACCGGATGGG 1875
 AAGCTGAGCTCAGGACTCCCCCAGCAGAAGCCCGCTGAGCAGGACACTCCAGTTCCTCTTCCAGCTGGCG 1950
 TCCTCAGAAGGAAAAGGCCCAGGAGGACCACCTGAAACACAGCATCAGCAGCCAGACTGTCTAGTGGGTCTGTC 2025
 GGGACCACCCAGACCACAGGAAGCAATACCGTCAGTGTTCAGCTACCTCAGCCCCAGCAGCAATACAACCATCCC 2100
 AACACTGCCACGGTTGCACCTTTCTGTACAGGTCTCAGACGGACACTGAAGGGAATGTTACTGCCGAGTCAAGC 2175
 TCAACAGGTGTGAGCGTGGAGCCCAGCCACTTCAACAAGACTGGACGCTGCTCTGGAAGAAGTCACTGCGGAA 2250
 GATCCTGAGGCTCGCGCGCTGCGGACAGTGAAGAAGCATCGCTGATCTGCGGCAGAATTTGGAGGAAACCATGTCC 2325
 AGTTTAAGGGGAAGTCAAGTTACACACAGCACATTGGAACCACGTTTGACACCAATGTCAACACGGAGATGAGT 2400
 GGCCGTAGCATACTCAGCTTGACAGGGAGGCCCCACACCTCTGTCTGGAGACTGGGGCAGTCCAGCCCTCGGCTC 2475
 CAAGCAGGAGACGCCCCCTCAATGGGCAATGGGTATCCCCCTCGAGCCAACGCCAGCAGGTTTCATCAACACTGAG 2550
 TCAGCTCGCTATGTGTACTCCGCCCCCTCTGAGAAGGCAGCTGGCCCTCCCGGGCAGTAGTGTCTGCCAYGTGGAC 2625
 GTCTCAGACAGTGAAGCAGGATGAGATGGAAGCTGGAAGGCATCAGCATGAGACGCCCGGCTACATGAGCGATGGG 2700
 GATGTTCTGAGCAAGAATCATCCGACCGATGACATTACAAGCGGATACATGACTGATGGTGGACTTGGCCTCTAT 2775
 ACCCGTCGCTGAACCGGCTCCCTGATGGGATGGCTGTGGTACGGGAGACCCTGCAACGAAATACCTCCCTGGGC 2850
 CTCGGAGACGCTGACAGCTGGGACGACAGCAGCTCCGTGAGCAGCGGCATCAGCGACACCATAGACAACCTCAGC 2925
 ACTGATGACATCAACACAGCTCCTCCATCAGCTCTTATGCCAACACACCTGCCTCTCTCGAAAAACCTGGAT 3000
 GTGCAGACTGATGCTGAGAAGCACTCAGGTGGAGGAAATCCCTGTGTTCTGGTGAATGTTCAAGAAATCA 3075
 GACGGAGGCTCAGACAGCGGCATAAAAATGGAGCCAGGTTCCAAGTGGAGGCGGAATCCTTCTGATGTGTCTGAC 3150
 GAKTCCGACAAAAGCAGCTCGGGCAAGAAGAATCCTGTCTCTCCAGACAGGCTCATGGCGCGAGGCATGACA 3225
 GCTCAGGTGGGCATCACCATGCCAAGGACGAAGGCTTCAGCCCCGGCAGGCGCACTGAAGACCCAGGAAGTGA 3300
 AAAACAGACGACGCAAGGTGTCTGAGAAAGGAAGGCTTCTCTAAAGCCTCCAGGTGAAGCGCTCCCCATCA 3375
 GATGAGGCGGAGCAGTGGTGACGAATCCAAAAGCCCCCTCCAGCAGCTCTAGGACACTACTGCCAATGCC 3450
 AACAGCTTTGGGTTCAAGAAGCAGAGTGGTTCCCGCGCGGCTGGCCATGATCAGCCAGCGGGGTGACTGTC 3525
 ACCAGCAGGTGAGCCACACTGGGCAAAATCCCAAAGTCATCTGCACTCGTCACTCGGTCTGCTGGTGGAGTCA 3600
 AGTATGGATGGGGCTCAGAATCAGGATGACGGGTATCTAGCCCTAAGCTCCCGGACAAACCTTCAGTACCGGAGT 3675
 TTGCCGAGGCCAGTAAGTCCAACAGCGGAACGGGGCTGGGGAACAGGTCTAGCACCAGCAGCATAGATTCCAAC 3750
 ATTAGCAGCAAGTCCGACGGCCTGCAGTGCCCAAAGTGGAGGACCTTCCAAAACAGCCCTAGGCACTCTCTA 3825
 CAGGCTTGGTCAACAAACAGATAAGGAGAAAGGATCTCTCAGACAAACGAGAGTGTGGCTTCTGTAACTCG 3900
 GTGAAAGTGAATCCGGCAGCCAGCCTGTGTCCAGTCCGGCTCAGACAGTCTCCAGCCTGGAGCCAAGTACCCA 3975
 GATGTGGCCTCTCCACACTCCGACAGCTCTTTGGTGGGAAGCCTACCAAGCAAGTGCCCATCGCCACAGCTGAA 4050
 AACATGAAAAATTCGGTGGTCACTTCCAATCTCATGCCACCATGACTCAGCAAGGTAACCTAGACTCCCCGTA 4125
 GGCAGTGGCGCTCTGAGCAGTGGGAGCAGCAGTCTCTCTACAGCAAGAATGTGGACCTCAACAGCTCTCCGTA 4200
 GCCTCCAGCCCCAGCTCAGCCCACTCGGCCCTTCCAACAGCCTCACTGGGGCAACCAAGCAGCAGCTCCTCC 4275
 GCAGTTAGCAAGGATGGCCTGGGCTTTCAGTCTGTGAGCAGCCTCCACACCAGCTGTGAGTCCATCGACATCTCC 4350
 CTCAGCAGTGGAGGGGTCCCCAGCCACAATTCTTCCACTGGCCTCATCGCCTCTCCAAGGACGACTCCTTGACT 4425
 CCCTTTGTGAGAACTAACAGTGTGAAGACCACACTGTGAGAAAGCCCTCTCTTCCCTGTCTAGCCCTAAG 4500
 TTCTGAGAACTACTCTGCCAGGAAACAGGACAGTGACCCGACCTTGATAGGAACACTTGCCTAAGAAAGGA 4575
 TCTAGGTAATCTCCACCTCCAGCTTCGACAGCAAGAAGATGCAAAAGAAATGGTTACGGTCCCATTTCTGCAGGA 4650
 GGCCTTCAGGACACCGCTGCCAATTCUCCCTTTTCTCTGGCTCCAGCGTGACTTCTCCCTCCGGAACAAGATT 4725
 AACTTTTCCAGCTTGGAGTCCCACCACTGTCAACAGATGAGCTTGTCCAACCCGACCATGCTGAGGACTCAC 4800
 AGCCTCTCCAATGCTGATGGGCAATGATCCATACACTGACAGCGCTTCCGGAATAGCTCCATGTCCCTGGAT 4875

5/56

Figure 1C (CONTINUED)

GAGAAGAGCAGAACCATGAGCCGTTTCAGGCTCATTCGGGATGGGTTTGAAGAAGTTTCATGGATCCTCACTCTCC 4950
 TTGGTTTCCAGCACATCGTCAGTTTATTCTACACCAGAAGAAAAATGCCAGTCAGAGATTTCGCAAGCTGCCGCCG 5025
 GAACGGATGCCCTCCAGGAGAAAGTTTCAGCTTTGACCACCCAGCTGACAGCAAAYGCTCACCTTGTGGCWC 5100
 TTTGAACAGAGTCTTGGTAAACATGACAATCAGGCTCCAGAGTCTGACCATGACAGCTGAGCAGAAGGATTCAGAA 5175
 CTGAATGAGTTAAGAAAAACCATGAGCTGCTAAAGAAACAGAACCGCAGCTGCCAGGCTGCCATTAAATGGAGTA 5250
 ATTAACACACCTGAGCTCAACTGCAAAGGAAACGGCACTGCCAGTCTGCAGACCTCCGCATCCGCAGGCAGCAC 5325
 TCCTCAGACAGCGTCTCCAGCATCAACAGTGCCACCAGCTCCAGTCTGGGGCAGCAACATAGAGAGTGA CTCA 5400
 AAGAAGAAGAAGGAAAGAACTGGgtcaatgagTTACGCAGCTCCTTCAAGCAAGCTTTCCGGGAAGAAGAAGTCC 5475
 CCAAAATCTGGCTCTCTCATTTCAGATATTGAGGAGATGACGGATTCTTCTTTGCCCTTCCTCACCAGTACCR 5550
 CACAATGGGTCCACAGGTTCCACCCCACTGCTGAGGAATTCTCACTCCAACCTCTCTAATTTTCMGAATGCATGGAT 5625
 AGTGAAGCTGAGACCGTCAATGCAGCTCCGAAATGAGTTAAGAGACAAGGAGATGAAGCTGACRGATATCCGCTTA 5700
 GAAGCTCTCAGTTCTGCCACCAGCTGGACCTCCGGGAGGCCATGAACAGGATGCAGAGTGAATAGAGAAG 5775
 CTGAAAGCTGAGAATGATCGGCTGAAGTCAGAGTCTCAAGGCAGTGGCTGCAGCCGGGCTCCTTCCCAAGTGTCC 5850
 ATCTCTGCTCTCCCGAGGCAGTCCATGGGCTCTCCAGCACAGCTTGAACCTCACTGAGTCAACCAGCCTGGAC 5925
 ATGTTGCTGGATGACACTGGTGAATGCTCGGCTCCGAAGGAAGGAGGCAGGCATGTTAAGATAGTTGTTCAGCTTT 6000
 CAGGAGGAAATGAAGTGAAGGAGGATTCAGACCAACAYCTCTTTCTTATTGGCTGCATTGGAGTTAGTGGCAAG 6075
 ACGAAGTGGGATGTGCTCGATGGGGTGGTTAGACGGCTGTTCAAAGAATACATCATTCATGTCCAGCCAGTGA 6150
 CAGCTAGGGCTGAATTCAGACAGCTGCTGGGCTACAGCATTTGAGAAATCAAGCGCAGCAACACTTCCGAAACA 6225
 CCGGAGCTGCTCTCTTCTGGCTATCTGCTTGAAGAGAACAGACCATCTCTCTCTCTGTAAGGCTCCGAGAA 6300
 AACAGCTTGGACTCACTGGTGTGTTGAGTCTTGTATTCCCAAGCCCATCCTGCAGCGCTACGTCTCCCTCCTGATA 6375
 GAGCACCGTCCGATCATTTCTCTCTGGCCCCAGCGGCACTGGGAAAACCTACCTGCCCAACCGGCTGTCTGAGTAT 6450
 ATAGTCTTTCGAGAGGGACGGGAGTTGACAGACGGGGTTATCGCCACCTTTAAGCTGGACCATAAGTCCAGCAAG 6525
 GAATTCGCCCAGTACCTGTCCAACCTTGCTGACCACTGCAACAGTGAGAACAATGCTGTGGACATGCCCTCGTC 6600
 ATCATCTTGGACAACTACACCACTGAGCTCTCTGGGCGAGATCTTCAATGGGCTGCTCAACTGCAAGTACCAC 6675
 AATGCCCTTACATAATTTGGCACATGAACCAGGCTACCTCTTCTGACTCCCAACCTGCAGCTTCACCATAACTTC 6750
 AGATGGGTGCTTTGTGCCAACACACGGAGCCTGTGAAGGGTTTCTTGGCCGATTCCTGAGGAGGAAGCTCATG 6825
 GAAACAGAGATCAGTGGGCGGGTCCGCAATATGGAGCTGGTAAAAATCATTGACTGGATTCCCAAGGTCTGGCAT 6900
 CACCTCAACCGCTTCTTGGAGGCTCACAGTTCTCTGGAGCTCACCATCGGCCCCCGGCTCTTCTGTCTATGCC 6975
 ATCGATGTGGACGGCTCCAGAGTGTGGTTCACCGACTTGTGGAATATTCCATTATCCCTATCTCTCTGGAAGCC 7050
 GTCAGAGAAGGACTCCAGCTCTATGGAAGGCGCGCCCCCTGGGAGGATCCTGCCAAGTGGGTGATGGACACATAT 7125
 CCATGGGCAGCCAGCCACACAGCAGGAGTGGGCTCCCTGTCTGCACTTACGGCCTGAGGATGTCTGGCTTCGAC 7200
 GGCTACTCCATGCCTCGGGAGGGATCGACAAGCAAGCAGATGCCCCCCAGTGATGTGTAAGGTGACCCGCTGATG 7275
 AACATGCTGATGAGGCTGCAGGAGGCAGCCAACTCTCAGCCCCCAGAGCTATGACAGCGACTCCAACAGCAAC 7350
 AGCCATCAGGATGACATCTTGGACTCCTCTTTGGAGTCCACTCTGTGACAGGGGCCCCGAGGCCAGCGCCCTCT 7425
 CTCTCTCTCACCAGATTCCACCTGCATCCCCACATCACCTTGAAGATGACTTCTTGAGCCAGCCCCAGCCACA 7500
 GCCTTAGAGCTGCGGGAACACCGAGACCCCCCGTCTTCAGCCTCGACCTGGGTGCAGGCATCCCGGGCCAGCTG 7575
 CCTGCGGACCGCTTCTTCCACAGCGAGAACTGCACCTTCTGTTGACTTAAATTATTGTTTTGCCCTTGTG 7650
 CTGTGACCTCCCTAAGACACTGAAGATACTTCTCGGGAAGGATCATCGCCGTTGAAATGAAAAA 7725
 AAAAAAAAAAAAAAAAAAAAAA 7748

At multiple positions heterozygous sequences have been observed. The ambiguities are denoted in the IUPAC IUB codes, which are as follows : R = A or G ; Y = C or T ; W = A or T ; M = C or A.

The region between position 5425 and 5433 is absent in cDNAs from Hela and colorectal adenocarcinoma tissue. Other cDNA sources are heterozygous (fragment present and absent) at this position.

cDNA from frontal cortex is heterozygous for the presence or absence of the region between 5924 and 6024. Absence of this fragment results in an out-of-frame deletion of 101 bp, resulting in a premature stop in translation.

The sequence in bold corresponds to the fragment in the 3'-UTR Hs-unc-53/2 that was used in RH mapping. The primers used to amplify SHGC-33456 are underlined.

6/56

Figure 1c (CONTINUED)

Three variants have been found for the 5' end of the gene. For these variants, the sequence from position 1 to position 366 should be replaced by one of the following sequences :

Variant 1
TGAAGAGGTGGTGGCTGATTTCCTTGGCTGGCGGGAAGCTCTGTCTGGCTGTTGCATGCATCACTTTTGTGTGGGTT 75
ATTTTGTTCCTCTGTGGATTGGAAGCATCGCTGAAGGAGAGAGAGGATTTTATTTCTGGGAAATGGAATCGGTT 150
TCTGAGTCCAGCCAACAGCAGAAGAGAAAGCCAGTTATCCACGGACTGGAAGATCAAAAGAGG 213

Variant 2
TGATACTTTGGGGTGCACATGGCTATTGATCTCTACTGCGGTTTGGCTTGTCTGTGGGGAATACATGAGCCCCGA 75

Variant 3
TAACAACTGGACTTTATTGAGTGTTTACCATGCACCAAGCCCTGGGCTAAACACTTCATCTGCAGGCTGTTCGTC 75
TTTACGGCAAACCCAGTAGGTAGGTATACTATCCCCACTCTGCAGATGCAGAAACGGAGGCACAGAGTGTTTTG 150
GTAGCTAAACAAGCTCACCAGGAGGCTAGAAGGTGGCCACACCTAGCTGGCCCCCTGACTCCACCAACTGCCTC 225
CCTTTGCTGTGTTGCATGCAAGAATGTGACTCCAAGTTTTTCCTTCCTTCTGGATCCAACCTCTGGCTTCACTCTG 300
CTCAGCAACCAG 312

7/56

Figure 1d. Amino acid sequence of the protein encoded by the Hs-unc-53/2 gene

```

mPAILVASKmKSGLPKPVHSAAPILHVPPARAGPQPCYLKLGSKVEVSKTTPSYQIPLKSQVLQGLQEPAGEGLP 75
LRKSGSVENGFDTOIYTDWANHYLTKSGHKRLIKDLQDDVTGVLQAIIQVANEKIEDINGCPKNRSQMIENI 150
DACLNFLAAKGINIQGLSAAEIRNGNLKAILGLFFSLSRKQOQQQQPQKQHLSSPLPPAVSQVAGAPSQCQAGTP 225
QQQVPVTPQAPCQPHQAPHQQSKAQAEQSRSLSGPTARVSAAGSEAKTRGGSTTANNRRSQSFNNYDKSKPVTS 300
PPPPSSSHEKEPLASSASSHPGMSDNAPASLESGSSSTPTNCSTSSAIPQPGAATKPWRSKSLSVKHSATVSMLS 375
VKPPGPPEAPRPTPEAMKPAENNQKSMLEKLLFNSKGGSKAGEGPGSRDTSERLETLPSEEESELEAASRMLT 450
TVGPASSSPKIALKGIAQRTFSRALTNKKSSSLKNEKEKEKQREKDKESKDLAKRASVTERLDLKEEPKEDPS 525
GAAVPEMPKKSSKIASFIPKGGKLNLSAKKEPMAPSHSGIPKPGMKSMGKSPSAPAPSKEGERSRSGKLSSGLPQ 600
QKPQLDGRHSSSSSSSLASSEGGKPGGTTLNHSISSQTVSGSVGTTQTGTSNTVSVQLPQPOQQQYNHPNTATVAPF 675
LYRSQDTEGNVTAESSSTGVSEPSHFTKTGQPALEELTGEDPEARLRVTVKNIAADLRQNLLEETMSSLRGTQVT 750
HSTLETTFTDNVTTEMSSGRSILSLTGRPTPLSWRLGQSSPRLQAGDAPSMGNGYPPRANASRFINTESGRYVYSA 825
PLRRQLASRGSSVCHVDVSDKAGDEMDELEGISMDAPGYMSDGDVLSKNIRTDITSGYMTDGGGLGLYTRRLNRLP 900
DGMVVRETLPQNTSLGLGDADSWDDSSSVSSGISDTIDNLSTDDINTSSSISSYANTPASSRKNLDVQTDAEKH 975
SQVERNLSLWGGDDVKSDDGSDSGIKMEPGSKWRRNPDSVSDXSDKSTSGKKNPVISQTGSWRRGMTAQVGITMP 1050
RTKASAPAGALKTPGTGKTDDAKVSEKGRLLSPKASQVKRSPSDAGRSSGDESKKPLPSSSRTPTANANSFGFKKQ 1125
SGSAAGLAMITASGVTVTSRATLGKIPKSSALVSRAGRKSSMDGAQNQDDGYLALSSRTNLQYRSLPRPSKSN 1200
SRNGAGNRSSTSSIDSNISSKSAGLPVVKLREPSKTALGSSSLPGLVNQTDKEKGISSDNESVASCNSVKVNPAAQ 1275
PVSSPAQTSLOPGAKYPDVASPTLRRLFGGKPTKQVPIATAENMKNSVVISNPHATMTQQGNLDSPSGSGVLSSG 1350
SSSPLYSKNVDLNQSPPLASSPSAHSAPNSLTWTGNASSSSAVSKDGLGFQSVSSLHTSCESIDISLSSGGVPS 1425
HNSSTGLIASSKDDSLTPFVRTNSVKTTLSSEPLSSPAASPFCRSTLPRKQSDPHLDNRNLTLPKKGLRYTPTSQ 1500
LRTQEDAKEWLRSHSAGGLQDTAANSFSSGSSVTSPSGTRFNFSQLASPTTVTQMSLSNPTMLRTHSLSNADGQ 1575
YDPYTDSRFRNSSMSLDEKSRMSRSGSFRDGFEEVHGSSLSLVSSTSSVYSTPEEKQSEIRKLRLRELDASQEK 1650
VSALTTQLTANAHLVAAFEQSLGNMTIRLQSLTMTAEQKDSLELNELRKTIELLKKQNAQAQAINGVINTPELNC 1725
KNGGTAQADLRIRRHSSDSVSSINSATSHSSVGSNIESDSKKKKRKNWVneLRSSFQAFGKKKSPKSASSHS 1800
DIEEMTDSSLPSSPKLPHNGSTGSTPLLRNHSNSLISECMDEAETVMQLRNLRLDKEMKLTDIRLEALSSAHQ 1875
LDQLREAMNRMQSEIEKLKAENDRLKSESQSGSCSRAPSQVVISASPRQSMGLSQHSLNLTESTSLDMLLDDTGE 1950
CSARKEGGRHVKIVVSFQEEWKWEDSRPHLFLIGCIGVSGTKWDVLDGVVRRLFKEYIIHVDPVSQGLNDS 2025
VLGYSIGEIKRSNTSETPELLPCGYLVGENTTISVTVKGLAENSLDSLVSFESLIPKPIQRYVSLLEHRRILS 2100
GPSGTGKTYLANRLSEYIVLREGRELTDGVIATFNVDHKSSKELRQYLSNLADQCNSENNAVDMPLVIIIDNLHH 2175
VSSLGEIFNGLLNCKYHKCPYIICTMNQATSSSTPNLQLHHNFRWVLCANHTEPVKGLGRFLRRKLMETEISGRV 2250
RNMELVKIIDWIPKVWHHLNRFLEAHSSSDVTIGPRLFLSCPIDVDGSRVWFTDLWNYSIIPYLLEAVREGQLY 2325
GRRAPWEDPAKWVMDTYPWAASPOQHEWPPLLQLRPEDVGFQGYSMFPREGSTSKQMPPSDAEGDPLMNMLRLQE 2400
AANYSSPQSYSDSDSNSNSHDDILDSSLESTL 2432

```

Putative start methanionines at positions 1 and 10 are in lower cases. The residue at position 1018 (denoted by x) is encoded by an heterozygous sequence. Both residues Aspartic acid (D) or Glutamic acid (E) can be incorporated. The amino acid sequence VNE at position 1776 to 1778 is present or absent depending on the allele from which the protein is translated.

For translation of the 3 variants described in figure 1c, the aminosequence from position 1 to 89 has to be replaced by the following amino acid sequences :

Variant 1	
mESVSESSQQQKRKPVIHGLEDQKR	25
Variant 2	
mAIDLYCGLACLWGIHEPr	19
Variant 3	
mQECDSKFFLPSGNSNGFTLLSNQ	24

8/56

Figure 1e. Nucleotide sequence of Hs-unc-53/3.

TAGAAGCATTTTCTTTGGCAGCAAGAAGATAATTTTATAGAAGCCATGCCTGTTCTTGGGGTTGCCTCAAAACTG 75
 AGGCAGCCAGCTGTTGGGTCAAAGCCTGTGCATACTGCTCTTCCGATACCAAATCTTGGCACTACTGGGTCACAG 150
 CACTGTTCTTCAAGACCTTTGGAACCTTGCTGAAACAGAGAGCTCCATGCTTTCTTGTAGCTTGCCTTAAATCA 225
 ACCTGTGAATTTGGAGAGAAGAAACCCCTCCAAGGAAAAGCCAAAGGAGAAAGAAGACAGCAAGATTTACACTGAC 300
 TGGGCCAACCACTACCTAGCAAAATCAGGCCACAAGCGGCTGATCAAGGACTTGCAACAAGACATTGCAGATGGA 375
 GTACTCCTAGCAGAAATCATCCAGATTATTGCAAAATGAAAAAGTTGAAGATATCAATGGATGTCTTAGAAGTCAG 450
 TCTCAGATGATTGAAAATGTTGATGTCTGCCTTAGTTTCTAGCAGCCAGAGGGGTAAATGTTCAAGGTCTATCT 525
 GCTGAAGAAAATAAGAAATGGAACCTTAAAGCCATTCTAGGGCTGTTTTTCAGTTTATCTCGCTACAAGCAGCAA 600
 CAACACCATCAACAACAGTACTATCAGTCCTTGGTGGAACTTCAGCAGCGAGTTACTCACGCTTCCCCTCCATCG 675
 GAAGCCAGCCAGGCCAAAACCCAGCAAGATATGCAGTCCAGTCTGGCAGCCAGATATGCAACTCAGTCTAATCAC 750
 AGTGAATTGCAACCAGTCAAAAAAGCCTACTAGGCTTCCAGGGCCCTCTAGGGTGCCTGCTGCAGGAAGCAGC 825
 AGCAAGGTCCAGGGAGCCTCTAATTTAAATAGGAGAAGTCAGAGCTTTAACAGCATTGACAAAAACAAGCCTCCA 900
 AATTATGCAAAATGGAAACGAAAAAGATTCTCTCAAAGGACCTCAATCGTCTTCAGGTGTAAATGGTAACGTGCAG 975
 CCTCCCAGTACTGCTGGGCAGCCTCCTGCCTCTGCCATCCCTTCTCCAAGTGCCAGCAAGCCCTGGCGCAGCAAG 1050
 TCCATGAATGTCAAACACAGTGCACCTCCACCATTGTGACTGTAAAGCAGTCAAGTACAGCCACCTCCCCACA 1125
 CCATCTTCAGACAGACTGAAGCCACCTGTCTCAGAAGGGTCAAAAGCTGCTCCCTCAGGACAGAAATCCATGCTT 1200
 GAGAAATTCAGCTAGTCAATGCCCCGACTGCTTACGCCCCCGCAGCCTCCAGTTCCAGGACCTAGTGATGGT 1275
 GGGAAAGGATGATGATGCCTTTTCTGAATCTGGTGAATGGAAGTTTAAACAGTGGTCTGAATAGTGGTGGCTCA 1350
 ACAAATAGCAGTCCCAAAGTGTCACCTAAGTTGGCCCCCTCCAAAAGCTGGAAGCAAAAATCTCAGCAATAAAAAG 1425
 TCTTTGCTACAGCCAAAGGAAAAAGAAAGAAAGAACAGGGACAAAAATAAAGTTTGCCTGAAAAACCACTCAA 1500
 GAAGAGAAGGATCAGGTGACAGAGATGGCTCCAAAAAGACCTCCAAAATTGCAAGCTTGATCCCTAAGGGCAGC 1575
 AAGACAACAGCAGCAGTAAGAAGGAAAGCTTAATTCGGTCTTCCAGTGGTATTCCAAAACCAAGGCTCTAAAGTTCCA 1650
 ACAGTAAAGCAAACCAATTTACCTGGCAGCACAGCAAGCAAGAGTCTGAGAAATTCAGGACTACCAAGGGGAGC 1725
 CCTTCCAGTCTCTTATCTAAGCCTATAACCATGGAGAAAGCAAGTGCTTCTAGTTGTCTGCCCCCTTGGGAAGGA 1800
 AGGGAAGCTGGCCAAAGCTTCTCCTTCTGGTCTCTGTACCATTGACAGTGGGCACAAAGCAGTGGCAGAGCAGGA 1875
 AATGTGCTGTGTCACCACTCCCTCAACAGCAGCAACATAGCCACCCGAATACCGCGACAGTGGCACCATTCTATTAC 1950
 AGGGCACATTCAGAAAATGAAGGTACCGCTTTACCATCGGCTGACTCCTGTACCAGTCTACAAAGATGGACTTA 2025
 TCATATAGTAAGACTGCTAAGCAGTGCCTGGAGGAGATATCTGGTGAAGGCCCTGAAACAAGAAGAATGAGAACA 2100
 GTTAAAAACATAGCAGACTTGAGGCAGAATTTAGAAGAGACTATGTCCAGTCTTCGTGGGACTCAGATAAGCCAC 2175
 AGCACCTGGAGACAACATTTGACAGCACTGTGACAACAGAGTTAATGGAAGGACCATACCCAACCTTGACAAGT 2250
 CGACCCACCCCATGACCTGGAGGTTGGGCCAGGCATGTCCGCGACTTCAGGCGGGAGATGCTCCCTCCCTGGGT 2325
 GCTGGCTATCCTCGCAGTGGTACCAGTTCGATTTCACACACAGACCCCTCGAGGTTCAATGATACACGCTCTC 2400
 CGTCGAGCTGCTGTCTCTAGGCTGGGAAACATGTACAGATTGACATGAGTGAGAAAGCAAGCAGTGACCTGGAC 2475
 ATGTCTTCTGAGGTGATGTGGGTGATATAGTGTGATGGTGATCTCTTGGGAAAAGTCTCAGGACTGATGAC 2550
 ATCAACAGTGGGTACATGACAGATGGAGGACTTAACCTATATACTAGAAGTCTGAACCGAATACCAGACACAGCA 2625
 ACTTCCCGGGACATCATCCAGAGAGGGGTTACAGATGTGACAGTGGATGCAGACAGCTGGGATGACAGCAGTTCA 2700
 GTGAGCAGTGGTCTCAGTGACACCTTGATAACATCAGCACTGATGACCTGAACACCACATCCTCTGTACGCTCT 2775
 TACTCCAACATCAGCTCCCTCTAGGAAGAATACTCAGCTGAGGACAGATTGAGAGAAACGCTCCACGACAGAC 2850
 GAGACCTGGGATAGTCTGAGGAACCTGAAAAAACAGAAAGATTTTGACAGCCATGGGGATGCTGGTGGCAAG 2925
 TGGAAGACTGTGTCTCTGGAATTCCTGAAGACCCGAGAAGGCAGGGCAGAAAGCTTCCCTGTCTGTTTCACAG 3000
 ACAGGTTCTGGAGAAGAGGCATGTCTGCCCAAGGAGGGGCGCCATCTAGGCAGAAAGCTGGAACAAGTGCACCTC 3075
 AAAACACCCGGGAAAAACCGATGATGCCAAAGCTTCTGAGAAAGGAAAAAGCTCCCTTAAAGGATCATCTCTACAA 3150
 AGATCTCTTTCAGATGCAGGAAAAAGCAGTGGAGTAGAGGAAAGCCCCCTCAGGCTATGGAATTCAGTCT 3225
 GCCACCACTCCTTTTGGCTTTAAGAAACCTAAGTGGAGTAGGGTCTATCTGCCATGATCACCAGCAGTGGAGCAACC 3300
 ATAACAAGTGGCTCTGCAACACTGGGTAAAATTTCCAAAATCTGTGCCATTGGCGGGAAGTCAAATGCAGGGAGA 3375
 AAAACCACTTTGGACGGTTTACAGAAATCAGGATGATGTTGTGTGCTGATGTTAGCTCAAAGACTACCCTACAATAT 3450
 CGCAGCTTGGCCCCGCCCTTCAAATCCAGCACCAGTGGCATCTCTGGCCGAGGAGGCCACAGATCCAGTACCAGC 3525
 AGTATTGATTTCAAACGTCAGCAGCAAGTCTGCTGGGGCCACCACCTCGAAACTGAGAGAACCAACTAAAATTGGG 3600
 TCAGGGCGCTCGAGTCTGTACCCGTCAACCAAAACAGACAAGGAAAAGGAAAAAGTAGCAGTCTCAGATTTCAGAA 3675
 AGTGTCTTCTTGTACAGTTCCCCCAAATCCAGCCCCACCTCTGCCAGCGCTGTGGTGCACAAGGTCTCAGGCAG 3750
 CCAGGATCCAAGTATCCAGATATTGCTCACCACATTTGGAAGGttgttgggtgccaagggcaggtggcaaatct 3825
 gctctgcacctaataactgaggggtgtgaaatcttctcagtaatgccagccctagtaccacatagcgcggcaaa 3900
 ggcagctctggagtcacgtcaggtcagggcagcatggcagctgtgtgggtgagcagcagcagccct 3975
 ctcttcaataaaacctcagacttaactacagatgttataagcttaagtcactcgttggcctccagcccagcatcg 4050
 gttcactctttcacatcaggtggtctcgtgtgggctgccaatatgagcagttcctctgcaggcagcaaggatact 4125
 ccgagctaccagttccatgactagcctccacacagagctcgtgagtcattgacctccccctcagccatcatggctcc 4200
 ttgtctggactgaccacagggcactcacaggtccagagcctgctcagtgagaacgggtagtgagatctactctc 4275
 tcagaaagctgacagcttgacagaaatcacactacccaaaaagggaactaagATATACCCCATCATCTCGGCAGGCC 4350
 AACCAAGAAGAGGGCAAAGAGTGGTTGCGTTCTCATTCTACTGGAGGGCTTCAGGACACTGGCAACCAGTCACCT 4425
 CTGGTTTCCCTTCTGCCATGTCTCTTCTGCAGCTGGAAAAATACCACTTTTCTAATCTGGTGAGCCCAACAAT 4500
 TTGCTCARTTTAACCTTCCCGGGCCAGCATGATCGCTCAAAAGCATCCCAAGCAAGACTCTTCTCTCGAT 4575
 CTCCTATGATGACTCCCAAGCTTTGTGGGAGTGCCACTTCTCTGGAGGAAAGACCTCGTGCCTCAGTCATTCTGGGC 4650
 TCATTACAGACAGCATGGAAGAAGTTTATGGCTCTTCAATTATCACTGGTGTCCAGCACTTCTTCTCTTACTCT 4725
 ACAGCTGAAGAAAAGGCTCATTACAGAGCAAATCCATAAAGTGGGAGAGAGCTGGTTGCATCACAAGAAAAAGTT 4800
 GCTACCCTCACATCTCAGCTTTTACGAAATGCTCACCTTGTAGCAGCTTTTGAAGAGAGCTTAGGGAATATGACT 4875

Figure 1e (CONTINUED) 9/56

GGCCGATTGCAAAGTCTAACTATGACAGCGGAACAAAAGGAATCTGAACTTATAGAACCTAAGAGAAACCATTGAA 4950
 ATGCTGAAGGCTCAGAATCTGCTGCCAGGCGGCTATTTCAGGGAGCACTGAATGGTCCAGACCATCTCCCAAA 5025
 GATCTTCGCATCAGAAGACAGCATTCCTCTGAAAGTGTTCCTAGTATCAACAGTGCACAAAGCCATTCCAGTATT 5100
 GGCAGTGGTAATGATGCCGACTCCAAGAAGAAGAAAAAGAAAACTGGGTGAactctagaggaaagtgaagtaga 5175
 AGTTCTTTCAAACAAGCCTTTGGGAAGAAAAAGTCCACCAAGCCTCTTCATCACATTCTGACATTGAAAGAGCTT 5250
 ACTGATTCATCCCTTCCGGCATCCCCAAGTTACCCATAATGCTGGTGAAGTGTGGCTCAGCATCCATGAAGCCC 5325
 TCACAATCTGCTTCAGCgtcacccttgtctggccaccaaaagacaaaatggccctgtgatctacaagcat 5400
 agatctcgGATCTGTGAATGCACAGAAGCTGAGGCAGAGATAATTCTGCAGCTGAAGAGCGAGCTCAGAGAAAAG 5475
 GAATTAATAATTAACGGATATTCGGCTGGAGGCCCTCAGCTCTGCTCATCATCTTGATCAGATCCGGGAAGCCATG 5550
 AACC GGATGCAGAATGAAATTGAAATACTGAAAGCTGAAAATGACCGGTTGAAGGCAGAACTGGTAACACAGCT 5625
 AAGCCTACTCGGCCACCGTCAGAATCCTCAAGCAGCACCTCCTCTTCATCTTCCAGGCAGTCATTAGGACTTTCT 5700
 CTAACAATTTGAACATCACAGAGGCTGTTAGCTCAGATATTTTGTCTAGATGATGCTGGTGAAGTGAACCTGGACAT 5775
 AAAGATGGCCGCGAGTGTGAAATTTATAGTCTCCATAAGCAAGGGTATGGTTCGAGCAAGGACAAAATCTCAG 5850
 GCATATTTGATAGGATCCATTGGTGTAGTGGGAAACCAAGTGGGATGCTTAGATGGTGAATAAGACGTCTC 5925
 TTTAAGGAATATGTATTCCGAATTGATACATCCACTAGCCTTGGTCTGAGCTCTGACTGCATTGCTAGCTACTGT 6000
 ATAGGAGACTTAATTAGATCCCATACCTAGAAGTGCCTGAATTGCTGCCTTGTGGATACCTTGTGGAGATAAT 6075
 AACATCACTACTGTGAACCTCAAAGGGGTAGAAGAAAATAGTTTGGACAGTTTTGTTTTTGATACGCTGATTCTCT 6150
 AAACCAATTACCCAAAGGTACTTTAACTTGTGATGGAGCATCAGAAATTATACTCTCAGGACCGAGTGGTACT 6225
 GGAAAGACCTATTTGGCAAACAACTTCTGTAATGTAATAACCAATCTGGAAGGAAAAAACAGAGGATGCA 6300
 ATTGGCCACTTTTAAATGTGGACCACAAGTCAAGTAAGGAATTGCAACAATATCTAGCTAACCTGGCTGAACAGTGC 6375
 AGTGTCTGATAATAATGGAGTGGAGCTCCAGTTGTAATAATTCTTGATAATCTTCATCATGTGGGCTCTCTGAGT 6450
 GATATCTCAATGGTTTTCTCAATTGTAATAACAACAAATGTCCATATATTATTGGAACAATGAATCAGGGAGTT 6525
 TCTTCATACCAAACTTAGAGCTGCATCACAATTTTCAGGTGGGTATTATGTGCAAAATCATAACAGACCTGAAA 6600
 GGCTTTTTTAGCGAGATATCTTCGAAGAAAACCTCATAGAGATAGAAATTGAAAGGAACATTGCAATAATGACCTA 6675
 GTCAAAATTTATAGATTGGATTCCGAAGACGTGGCATCATCTCAACAGTTTTTTGGAAACACACAGTTCTTCTGAC 6750
 GTTACCATTGGTCCCCGACTATTCTCTCTTGGCCCCATGGATGTAGAAGGTTCTAGAGTATGGTTTCATGGATCTC 6825
 TGGAACTATTCTTTAGTACCTTATATTCTGGAGGCAGTGAAGAGGGTCTTCAGATGTATGGGAAACGCACACCA 6900
 TGGGAAGATTTCTTTAGTACCTTATATTCTGGAGGCAGTGAAGAGGGTCTTCAGATGTATGGGAAACGCACACCA 6975
 TTACTTCAGCTGCGACCAGAAGATGTTGGGTATGAAAGCTGCACATCCACTAAGGAAGCCACAACCTCAAAGCAC 7050
 ATTCCACAACTGACACAGAAGGAGATCCCTGATGAATATGCTAATGAACTCCAAGAAGCAGCCAATTACTCG 7125
 AGCACACAAAGCTGCGACAGCGAAAGCACCAGCCACCATGAAGACATTTTGATTCTCTCTTGAATCTACCCCTC 7200
 TAGAGGGTGAAAAAGTTAAGGGAAGAGACTTTGCTTTTAAAAAAATGTTTCAAAAGAAAGGTATTTTCACTAAA 7275
 CCCTGCCAGTATAAAAGCACCTCTGCAAGGGCCCTGACCCAGAGTTGTGGTCTCCAAGGAGGCAGCAGAACTAA 7350
 GTCTGAACCGCCAAGATGCTAAATTGCAATGGAAGCTTAACTTTAGTTTATTTCTAAGCATTTTTTTATATCTGTG 7425
 GAGTAATAGAAAGCTCCATTACTCAACTGGAAGGACCCCTAATGACAGGGCAACTGAACAGATTGCACATGGGAT 7500
 AGCCAACTGGACTTTCTTTGTTTCTCTTTTAAAGCTTTTACAATGTCAGACCATTTTTTGTCCCTTCTTTTGT 7575
 CCTCTGAGGGGCTGTTTCGCCCCAGGCAGGGTCACTTTCTGATCTGTCCAACCTCCTTTGTGCCACACGGTGCT 7650
 GGTACAGGGGCTTCAGTAGTGTGTTGTGTTGTGCGCTCACCCCATTCAGAACAAATCCAAGAGGCCAGTCCCTCA 7725
 TAAGCACAAATGGAATTGTGCAACCACCAGAAAAACACTACTGTGGCAACTGGAGAAGTGCCAATTTAATTCTA 7800
 ACTGCCACGTTCTCATGATGTGCTCCACCACTTTTATGATATGAGTCACTGGTTTATAAGGTTGTTTTTACC 7875
 ACAGTGGTCTTTTTAAACCACTGCCACTCCCTTAAACAAGATTTTATACCAATTATTAGTCAACACTGATAAA 7950
 AGGCTTTTTTTAGGGCTTATTGTTTGTGAGCCTTTTCAGTGAAGAAGGAACATTTCTATGGTGTGCTCTCACTG 8025
 CCTTAAACAGATTTCTATGACAGTTTAAAGTGTGTTTAAATCCTAAACCATGGTAATTTCCACTGTCTTTTC 8100
 ATTTACAACCAAGCAACACCAGTTAACATAGTAGCCTCATCTCTATATATCTTTCTCTTTTTTTTTTTTGAAG 8175
 AAATGGATAGGAGAAAGATCAGTATTTTACCTTGTGATGATCGCTTGGCTATCCTCCAAAATATTAAAAAT 8250
 AACCAGAAATGCTCTTTGACCGTCACTTAAACCTTAAGACATGTGGCGAAATTCATCCAGTTCTAAGTGAAG 8325
 AGTTTCAGAAGGCAGGAGATTTTGAATTATTATCCAGCAGGGCTGGAAGCACTAGATGCAGCATGAGCACAACCTA 8400
 TTCGGCTTTCCTTCCCTATTGTTTTTGTGTTTTTAAAGTGTGTTGACGCATGTTGTTTTGATTGCTATTGTTGTA 8475
 CATGAGAAATTCAGCATTAAGAACACTGAAGCGGTAAAGTCACTGTGGAAGAGGAAGCGTTTATAGGTGAAAAAG 8550
 AAGGTTAGATTGTCACAGTCTACTGGGTAGGTATTGTAATAATAATTTTTTAAACTTGCACAAATCAAAACAAA 8625
 CACAAACAAAATTTGATTTTATCCTGTTGGTGTAAAGAGGTGTTTCACTTGTGAGATTCTCTGTACATTGCAAA 8700
 CAAATACAGAATGCAACCCCTCAAAGCTGTATTATCTGGTGTGTTTGTCTGTATTACAGTTGTTTTTGTACTAT 8775
 GCAGGAGCTATCAGTGTAGAGTGTGAGCATGCTTCAAACCTGTACATGAAGCCAATATATTTTTGGATAAGTAAA 8850
 AAAAAAAAAAAAAAATCGAGGGGGGGCCCGGTACCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAGTGGCC 8925
 GTCGTTTTACAACGTCGTGACTGG 8949

The region from position 3795 to 4325 consists of two blocks (3795 to 4283 and from 4284 to 4325) that independently can be present or absent in cDNA molecules from frontal cortex tissue. Frontal cortex is also heterozygous for the region from 5153 to 5173. The region from 5343 to 5408 is absent in frontal cortex, but heterozygously present in hart cDNAs.

The nucleotide sequence in heterozygous at position 4509. R is the IUB IUPAC code for A or G. Amino acid sequence is not affected.

An alternative 5' end has been observed. In this variant the sequence from position 1 to 288 is replaced by the following DNA sequence :

TAGTTTGTGCTTTTTTGAAGAGATTCCATTTTGAAGGGCAAGAACCCTAATGTGATGGATTATCTTCAGAAATG 75
 AACAGACATGGGAAGAATCCAGTGAGTCACAAGCTAGAAGATCAGAAGAAG

10/56

Figure 1f. Protein sequence encoded by the Hs-unc-53/3 gene

```

mPVLGVASKLRQPAVGSKPVHTALPIPNLGTGSGHQHSSSRPLELAETESSmLSCQLALKSTCEFGKKPLOGKAK 75
EKEDSKIYTDWANHYLAKSGHKRLIKDLQDDIADGVLLAEIIQIIANKEKVEDINGCPRSQSQMIENVDVCLSF 150
ARGVNVQGLSAEEIRNGNLKAILGLFFSLSRKQHQHQQYQSLVELQQRVTHASPPSEASQAKTQQDMQSSL 225
AARYATQSNHSGIATSQKKPTRLPGPSRVPAAGSSSKVQGASNLNRRSQSFNSIDKNKPPNYANGNEKDSSKGPQ 300
SSSGVNGNVQPPSTAGQPPASAI PPSASKPWRSKSMNVKHSATSTMLTVKQSSSTATSPTPSSDRLPKPPVSEGVK 375
TAPSGQKSMLEKFKLVNARTALRPPQPPSSGSPSDGGKDDDAFSESSEMEGFNSGLNSGGSTNSSPKVSPKLAPPK 450
AGSKNLSNKKSLLOPKEKEEKNRDNKVCTEKPVKEEKDQVTEMAPKKTSKIASLIPKGSKTTAAKKESLIPSS 525
GIPKPGSKVPTVKQTIISPGSTASKESEKFRITTKGSPSQSLSKPITMEKASASSCPAPLEGREAGQASPSGSC 600
VAQSSGQSTGNGAVQLPQQQHHSHPNATVAPFIYRAHSENEGTLPSADCTSPTKMDLSYSKTAQCLEEISG 675
EGPETRRMRTVKNIADLRQNLLETMSLRTGQISHSTLETTFDSTVTTEVNGRTIPNLTSRPTPMTWRLGQACPR 750
LQAGDAPSLGAGYPRSGTSRFIHTDPSRFMYTTPLRRAAVSRLGNMSQIDMSEKASSDLMSSEVDVGGYMSDGD 825
ILGKSLRTDDINSGYMTDGGNLNLYTRSLNRIPDTATSRDIIQRGVHDVTVDADSWDDSSSVSSGLSDTLNISTD 900
DLNTTSSVSSSYNITVPSRKNTQLRTDSEKRSTTDETDWDSPEELKKPEEDFDSHG DAGGKWKTVSSGLPEDPEKA 975
GQKASLSVSQTGSWRRGMSAQGGAPSRQKAGTSALKTPGKTDDAKASEKGKAPLKGSSLQSRSPSDAGKSSGDEGK 1050
KPPSGIGRSTATSSFGFKKPSGVGSSAMITSSGATITSGSATLGKIPKSAAI GGKSNAGRKTSLDGSQNDQDVVL 1125
HVSSKTTLQYRSLPRPSKSSSTSGIPGRGGHRSSTSSIDSNVSSKSAGATTSKLREPTKIGSGRSSPVTVNQTDKE 1200
KEKVAVSDSESVSLSGSPKSSPTASACGAQGLRQPGSKYPDIASPTFRRLfgakaggksasapntegvksssvm 1275
pspsttlarqgslespsstgsmgsagglsgsssplfnkpsdlttdvislshsllasspasvhsftsgglvwaanm 1350
ssssagskdtpsyqsmstslhtssesidplshhgsllsglttgthevqslmrtgsvrstlsesmqldrntlpkkg 1425
LrYTPSSRQANQEEGKEWLRSHSTGGLQDTGNQSPVSPSAMSSSAAGKYHFSNLVSPTNLSQFNLPGPSMMRSN 1500
SIPAQDSSFDLYDDSQLCGSATSLEERPRASHSGSFRDSMEEVHGSSLSLVSTSSLYSTAEKHAHQEIHKLR 1575
RELVASQEKVATLTSQLSANHLVAAFEKSLGNMTGRLQSLTMTAEQKESELIETIEMLKQNSAAQAAIQG 1650
ALNGPDHPPKDLRIRRHSSSESVSINSATSHSSIGSGNDADSKKKKKKNWVnsrgselRSSFKQAFGKKKSTKP 1725
PSSHSDIEELTDSSLPASPKLPHNAGDCGSASMKPSQSASAsplwppkkrqngpviykhrrICECTEAEAEII 1800
LQLKSELREKELKLTDIRLEALSSAHLDDQIREAMNRMQNEIEILKAENDRLKAETGNTAKPTRPPSESSSSTSS 1875
SSSRQSLGLSLNNLNITEAVSSDILLDDAGDATGHKDGSRVKIIVSISKGYGRAKDQKSQAYLIGSIGVSGKTKW 1950
DVLDGVIRRLFKKEYVFRIDTSTSLGLSSDCIASYICIGDLIRSHNLEVPPELLPCGYLVGDNNIITVNLKGVENS 2025
DSFVFDTLIPKIPITQRYFNLLMEHHRIILSGPSGTGKTYLANKLAEYVITKSGRKKTEDAIATFNVDHKSSKELQ 2100
QYLANLAEQCSADNNGVELPVVILDLNLHHVGLSDIFNGFLNCKYKCPYIIGTMNQGVSSSPNLELHHNFRWV 2175
LCANHTPEPVKGFGLGRYLRRKLIETIERNIRNNDLVKIIDWIPKTHHLNSFLETHSSSDVTIGPRLFLPCPMDV 2250
EGSRVWFMDLWNSLVPIYLEAVREGLQMYGKRTWPEDPSKWVLDTPWSSATLPQESPALLQLRPEDVGYESCT 2325
STKEATTSKHIPQTDTEGDPLMNLMLKQEAANYSSSTQSCDSESTSHHEDILDSSLESTL 2385

```

Regions corresponding to heterozygous sequences encoding presence or absence of this region are in lower case letters. These regions are from 1326 to 1413 ; from 1414 to 1427 ; from 1703 to 1709 and from 1768 to 1788.

Putative start methionines at positions 1 and 51 are indicated in lower case.

For the variant mentioned in figure 1e, the amino acid sequence from position 1 to 81 has to be replaced by the following amino acid sequence :

mDLSSEmNRHGKNPVSHKLEDQKK

24

11/56

Figure 1g. Nucleotide sequence of a 4984 bp fragment from BAC 585E09 (contains part of the genomic sequence of Hs-unc-53/1) extending the sequence derived from cDNA libraries shown in figure 1a.

```

TTCCTGATCTCAAGAGTTACTCCTTCCCCCTACAAAGCCCTCAGCCCCCTCCCCAGTCAACGCTAGGCCCTTCTC 75
TCCAAGCCACCGTGTCTTACCCCCATCCCCCTACCTCCTGGGCTCAGGAGGCAACCTTGAGCCTCAGAGACTGA 150
AGTAGGGTGGGACTGGGAGTTTCTGGGGGAAAACCAAAGACGGTTTGGGGTTGGGGAGGGGAATGAGCACCTT 225
GGATACCATCTCTCCACCCCTCTCCCGACATCTCTCTCAGGCCACAGGCCCACTTTCCCTCCCGCATTCTGAGC 300
CGCCCTCCCTCCGTCTCTTTTACCTGCACCTCCACACCTCTCAACAGATCTTTATCTGGACACGGCAGGGGGT 375
CCCCGTGCCCCCTCGAGAATCCAAGAACCCTCCCGCTTCTACGCGGAAAGCTGGGAGAAAACTGCTTTTCCTTT 450
ATTTCCCCCTACCCCCACTCATCGCCCCCTGGAGCTCCGCTCGCAGATACCTCCCCCTCCCGAGCCAGAAATAG 525
ACACACTATCTCTCCCCACCTCCCTCCCGGTGCGCACACTCGCTCCCTCTCTCTGTTTGTCTCCCGCTTCCC 600
CTTCCCTCTCTCTCTGCTCGGAGCTGCAGCTGCAGCTCGACTCGGGCTGGCTGGCTGGCTGAGTGGCGCCGG 675
GGCGCTGCCCCGCGAGTGGCGGTGTCCACGGGACTGACAGGCAGGCAGGCAGGCCGCGGGCTGGGATCCGGACACCA 750
AAGCAAAAGCAGCCGTGGGCGCGGAGGAGCCGCGGGCTTCCATCCTTCTTTGACTGATTTTTTAAATTTTAAT 825
TTGTATTTTCCCCGCGCCCCGCCCCCTTTTCTCCGACCCCGCCCTATCGCTCCCCGGCTTCCCTGCTCTTTCT 900
TTTTCCCGGCTTCTTCTCTCGCGTTTCTTTCCCTTGCGCCCTCGGCTTGCTCTCTCTCTCTCTCTCTCTCT 975
CCCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1050
CGCTCTCCCGCCCCCTGCCCCCTCCCCCGCTGCTCGACGCGCGGATCGTCCATGCGCTCTCGCGCGAGAAT 1125
GCTGGGCGAGCAGCTCAAGAGCGTGCAGCCCGAGCTGGAGCTGAGCAGCGCGCGCGCGCAGCAGGGCGCGGACGA 1200
ACCGCGGGGCGCGCGCAGGAAGGCGGCGAGCGCGGACGGCAGAGGCATGCTGCCAAGCGCGCCAAGGCGCCGG 1275
CGGCGGCGCGCGCATGGCCAAGGCCAGCGCGGCTGAGCTGAAGGTCTTCAAGTCCGGCAGCGTGGACAGCGTGT 1350
CCCCGGCGGGCGCGCCGCTTCAACCTGCGCAAGCAGAAGTCACTACCAACCTCTCTTTTCTACGGACTCCGA 1425
GAAAAAGCTGACGCTTTATGAGCCGAATGGAGCGACGATATGGCCAAGGCGCCAAAGGCTTAGGCAAGGTGGG 1500
GTCCAAGGGCGGTGAAGCTCCGCTGATGTCCAAGACGCTGTCCAAGTCCGAGCACTCGCTCTTCCAGGCCAAGGG 1575
CAGCCCGGCGGGCGGTGCCAAGACCCCTTGGCTCCGCTCGCGGCCAACCTGGGAAAGCCGAGCCGGATCCCTCG 1650
AGGACCTATGCGGAGGTCAAGCCGCTCAGCAAGGCGCCTGAAGCGGCGGTGAGCGAAGATGGCAATCGGACGA 1725
CGAGCTGCTCTCCAGCAAGGCCAAGGCGCAAAAGAGCTCTGGCCCTGTCCCTCTGCCAAGGGCCAGGAGGAGCG 1800
CGCCTTCCCTCAAGGTGGACCCCGAGCTGGTGGTGGTACCGTGTCTGGGAGACCTGGAGCAGCTGCTCTTACGCCAGAT 1875
GCTGGGTAAGTCTTCCGCCCCCGCCCCGCCCCGCCCCCTGGCTTTCTCTTAACCAGCTGCTGGGGAAGGTGTGGGG 1950
AAAGCGAAGCCCCCTCCCCCTTGGCGCTTCCCGGAGGCGCCCTCTGTTCACGATCAGGCTGTGATGGGCATTGCGC 2025
CCAGATGTGCTGAGCTGGCCACCTCCAGATGCGCATGGCTCAAGTGTACCTTCTTAAAGACATTACAGCGCGGA 2100
ACCGGGCTCGACTCTGGCTTGGCCGTGCCGAGGTGAGCTTGGGCTGTCCCTCTGCCAAGGGCCAGGAGGAGCG 2175
GAAATAGAGCCAGAATCCCAATATGGCAAAACCTGGGACTGGTGGAAACCTCCGTTGTGGTGTGGCCTGCGCTTG 2250
ACAGGAGCATCCCGCATTGCAAGGGGAGCGTCCAGCGAGAGCCCGGATCTAGAGGACAGATGTGGGAGAGCAGAT 2325
GTGAGGGCTGATTGGCCCCGGAACACAGCTGAGGCTCCACTTCTCTGTGGATCCCGAGTGGGAGCGCAAGTCCG 2400
ATTTCCCCCGGTGTGAGGATTCTGGCTAAAAGAGCGCTTAGGGCCGGGGCGGGCGGGCGGGCTGTGCGC 2475
ATCTGGGCGCATGTCCGATACCTCAGCCCCGGCTCTGGCCCCAACCCCTACACCCGAGGTCTTTTAGGGCGTGT 2550
CGAAAGCTCTGGGCGTTAGCGCCGAGACTCTGTTTGACCGGGAAGCCTTGGCCCTGTGGCTAATGGAAGCCGAG 2625
CAGGCGGGAAGGGAGGAACAAAGCTTGCTCGAGTGGAGGAAGCGCGCAGAGCTGTTCCATTGTTCTCCGTGCC 2700
AAGAGTCTATGCAAAAAAACCCGAAGCGGGCCCGGAACCTGCTCTTTCTCTCCCGGAGAGCCCTCGCCCTAGA 2775
GAGGAAATAGATCTGGGATGTGCGGAGCGCCAGGCGGAGCTGCTCGGGAACCTGCGACGGGCCCTTGGGCGCAC 2850
GGAACAAGGACGGTGGGGCTGGTGCCAGGCGAGCTGCTTTGGCTGCGCGGACTGTTGCGGTGGCTGGGTTGT 2925
GGGTCTTCCGGCGCCGAGGGACCCGAGCTTCTTGGGTACCCGGCAGGCTGCCCGCCCGCTGGGGGCTGGGAAGGG 3000
GCGTGCCAAATGCGCGTGTGAAAGGGCGGGGCGGAGTGACGGGCTTGGTGGGTGGGGAACATGCAAGAGCTCGCC 3075
GGCGGCTTGGGAATGCGAAGCCGAGGAGACCCGCTGAGGCTTGGCTGAGCTTCTCTAGGACCTTCCATCTCT 3150
GTGTGGGCTAGGATGAGGGTCTCTGACAGGGGCAAGGATTTGGGCTTTGGGAGAACCGATCTTACGCAGGAGG 3225
CCGCAATGGGCTTTGACAGGGCAATCAGGAGACTGGACAAGGGCAAAAGAAGAGCAGCCTTTTCCCTGGGAGC 3300
CCCTCTGAAGGTGGGGATGGCTGGGTGGGTGCGGAAGCTGACCAGGCAGCCTCACTCTGCAAGGGAATGTGCC 3375
ACCCGCTCTCAGTGTGGGGCTGAGCCTGTCAAAGGCCCTGCCTCAGTGAATGGGGCAAGAGAGACAATAAGGGA 3450
AAAAATTAATAAATTTTGGCAGGCACCATGGCAGGCAACCAAGGAGGATATGGACAAATGCAACTGGCCCATG 3525
TGATAGAGAGCCCTGCTGAGGGACTGAAAGCAGGGTGAGAGAGGAAGGGACCGTGTGTGTGTGTGTGTGTGTG 3600
TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 3675
GCACAGAATAGCAGAAGGCACAGAACCCTTTTACGGGTCAAGGCTTTACTTGTGGGATAACTTAGCTGGTCTG 3750
GGTCTCTCCAGACTGGATGCCCTCACACTGTGCCAGAGCTGAGTGGGATGAGTGAAGCCCTCTTAGTTGCTCT 3825
CAAGAGGAGCCAAACAGTCTGAGCTGGCTAGGGAGATGGGAGGAGGGGAGGAGTGGGAGGAGGGCAGGTGCA 3900
GGGAGGGCGAGAGAGGAGGAGAAGCTGAGCTGTGGTCCCTTATTCCTGCTTAGCAGTGTGCACTTCTCAAAGCAC 3975
ACTGACACTTTTCACTAACCTCGGAAGTGAGGAGAGAACCTCCACTTCCAGTTGGGGAAATGCAGAGTCAAAA 4050
GCATTGAGGGCTTAAAGGCATCTATGAGTTCTAGTGGTGAAGGGAGAGATCCACATTGATCTCCAGGACTAAGT 4125
CGCAGCTTCTTCCAGAGACCTGATTGAGAGAGGAAGTGTGACAGGCAGAGGACCTGCCCAAGGCTATGTCTA 4200
CTGGGTATGGGCCACCAGAAGTGCCTCGATGACCCTACAGAGGGCTGAGGGGCTTAGCTCTCTGGGGTGGGGAGA 4275
GAAGGGTGGAACCTCCCAATCTGCCTGTCTCCAGCTGAGAGGACCCAAAGTTGGGGGGTGGGGAGTTGGTTACG 4350
GCTGTAGCAAGGCAGAGCCTGGTGTCAAACAGTGGTAGGGAGGAAAGGAGGGAGTTGGTGACCTCAAACCTAAG 4425
CTTTTCCCTGTGTGAAGGGCAGAGGGTAGACTGCCTGGGGAGGGGTAGAGGGAGAGGAACACAGAGGGAATTC 4500
GTCTTCCAGAGCCAATGATGGTGGTGTTCAGGTATCAGACAGGCCCTCAGTGTACAGCAGGGTGGCCTCTGGGGA 4575
GAAGAATGGTGACTTGATGTTTCAAGATTGTGATTGAAGACACTGGGCATTTGTCCCCACCTCAGTGGGGCTCAG 4650

```

12/56

Figure 1g (CONTINUED)

```
TGTCCAGTTATGTTCACTCCATAGTACCATCCTAGATCCAAGAGGCTGCCAAGAATCAATTTCTGAGGCGGAGGG 4725
AGGGGGTGGGAGTGAGGCAGCTTCAAGTCAGAGCCTTTCTGTAATAAGAGGGAAGGACTGAAACCTGATCATCCC 4800
CTTCCCAGAAATCAGCTGGGGTCCCAGATGGTCTAGGCAGGCTCCCTGTCCCTTCGCTAACCTTGAAGCTGCCA 4875
AATAACTAGGGCCCCACTGGGGAACCCTAGCAACTTGAAGACTGAGGAGTGAGTACCGAGGGCAAATGGGCTAA 4950
TTCCAGGAATTAGATGCCTCTGGACCCTGGCCCCG 4984
```

The sequence shown in figure 1a starts at position 1246. Upstream in the same reading frame as used for the translation of the DNA sequence in fig 1a into the protein sequence of fig 1b, a stop codon is found at position 815. A first putative start codon (ATG) can be found at position 1124. Assuming this start codon, the protein sequence from fig 1b is extended by the sequence
MLGSSVKSQPEVELSSGGGDEGADEPRGAGRKAAAADGRG

Intronic sequence has been found to start at position 1881.

13/56

Figure 1h. Illustration of a 5'-deletion variant of Hs-unc-53/3 discovered by Nagase et al., (1999, DNA Res. 6:63-70).

>KIAA0938 protein, amino acid sequence
 MCVTKKLFIVQRTIFVGCVIWKFLHYVLRGFLCFNSMQLDRNT
 LPKKGLRYTPSSRQANQEEGKEWLRSHTGGLODTGNQSPLVSPSAMSSSAAGKYHFSN
 LVSPTNLSQFNLPGPSMMRSNSIPAQDSSFGLYDDSQLCGSATSLEERPRAISHSGSFR
 DSMEEVHGSSLSLVSTSSLYSTAEKAHSEIHKLRRELVASQEKVATLTSQLSANAH
 LVAAFEKSLGNMTGRQLSLTMTAEQKESELIETIEMLKQNSAAQAAIQGALNGPD
 HPPKDLRIRRHSSSESVSSINSATSHSSIGSGNDADSKKKKKKNWVNSRGSELRSFQKQ
 AFGKKKSTKPPSSHSDIEELTDSSLPASPKLPHNAGDCGSASMKPSQSASAICECTEAE
 AEIILQLKSELREKELKLTDIRLEALSSAHLDDQIREAMNRMQNEIEILKAENDRLKAE
 TGNTAKPTRPPSESSSSSTSSSSSRQSLGLSLNLLNITEAVSSDILLDDAGDATGHKDG
 SVKIIIVSISKYGRAKDQKSQAYLIGSIGVSGTKWDVLDGVIRRLFKEYVFRIDTSTS
 LGLSSDCIASYICIGDLIRSHNLEVPPELLPCGYLVGDNNIITVNLKGVEENSLDSFVFD
 LIPKPITQRYFNLLMEHHRIILSGPSGTGKTYLANKLAEYVITKSGRKKTEDAIATFNV
 DHKSSKELQOYLANLAEQCSADNNGVELPVVILDLNHHVGSLSDFINGFLNCKYNKCP
 YIIIGTMNQGVSSPNLELHHNFRWVLCANHTPEVKGFLGRYLRRKLEIEIERNIRNND
 LVKIIDWIPKTIWHHLNSFLETHSSSDVTIGPRLFLPCMDVEGSRVWFMDLWNYSLVPY
 ILEAVREGLQMYKRTWPWEDPSKWVLDITYPWSSATLPQESPALLQLRPEDVGYESCTST
 KEATTSKHIPQTDTEGDPLMMLMKLQEAANYSSSTQSCDSESTSHHEDILDSSLESTL"

>AB023155 cDNA nucleotide sequence

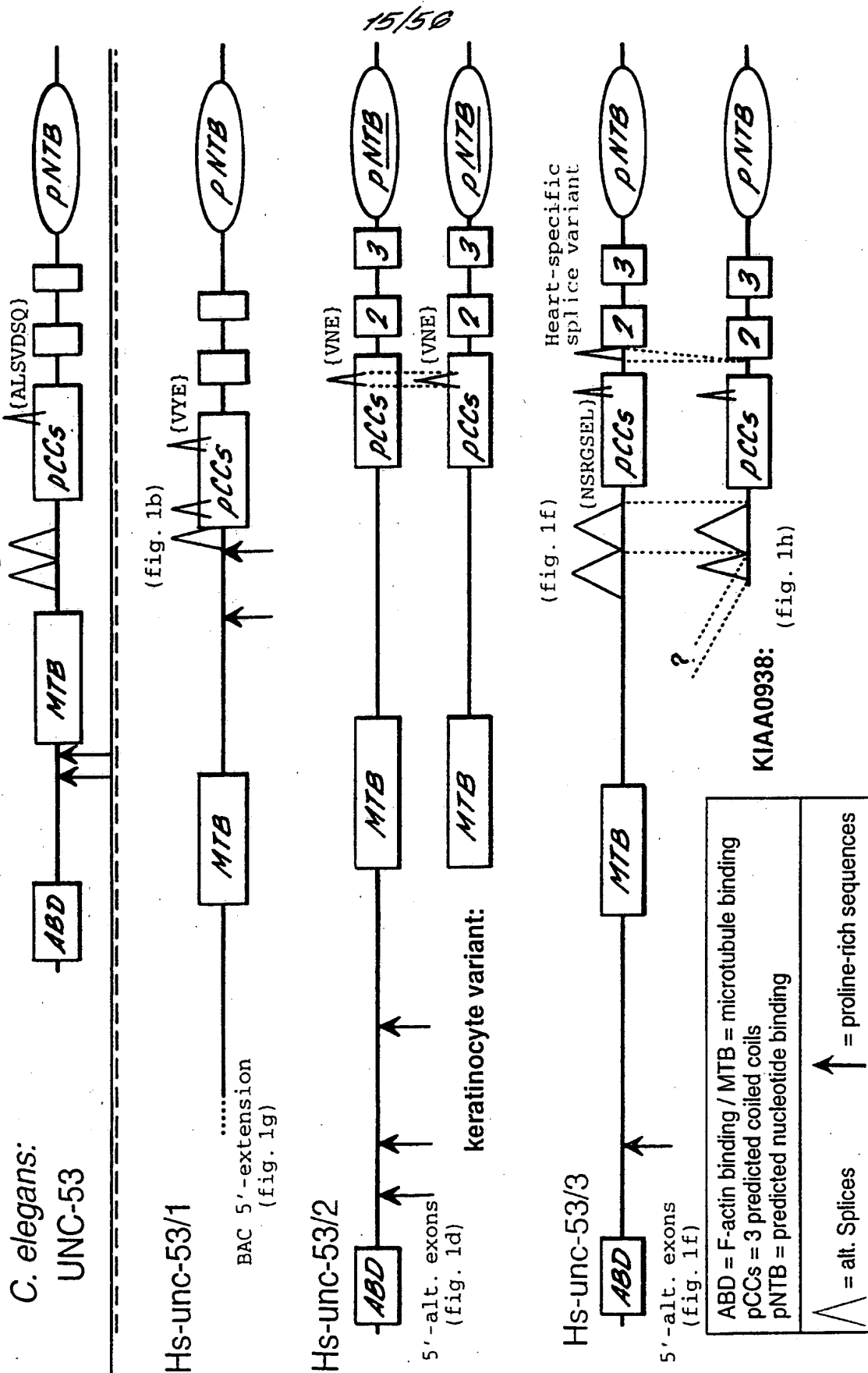
ctatcactaa	actgtcattg	aattgtactg	cattagaaag	gaactcaaat	atgtgtgacg	60
gcaatggaca	tcttgtcacc	tttagttggc	ctttttcaat	gagttaagca	ttatatgtgt	120
gttaccaaaa	aattattttt	tatagttcag	agaaccattt	ttgttgatg	tgtaatttgg	180
aagttttgtt	tacattatgt	ccttaggggt	tttctttgtt	ttaacagcat	gcagcttgac	240
agaaatacac	tacccaaaaa	gggactaaga	tatacccatc	catctcgga	ggccaacca	300
gaagagggca	aagagtgggt	gcgttctcat	tctactggag	ggcttcagga	cactggcaac	360
cagtcacctc	tggtttcccc	ttctgccatg	tcattctctg	cagctggaaa	ataccacttt	420
tctaacttgg	tgagcccaac	aaatttgtct	caatttaacc	ttccggggcc	cagcatgatg	480
cgctcaaaac	gcattcccagc	ccaagactct	tccttcgatc	tctatgatga	ctcccgactt	540
tgtgggagtg	ccacttctct	ggaggaaaga	cctcgtgcca	tcagtcattc	gggctcattc	600
agagacagca	tggagaaggt	tcattgctct	tcattatcac	tggtgtccag	cacttcttct	660
ctttactcta	cagctgaaga	aaaggctcat	tcagagcaaa	tccataaact	gcgagagag	720
ctggttgcac	cacaagaaaa	agttgtctac	ctcacatctc	agctttcagc	aaatgctcac	780
ctttagtagc	cttttgaaaa	gagcttaggg	aatatgactg	gccgattgca	aagcttaact	840
atgacagcgg	aacaaaagga	atctgaactt	atagaactaa	gagaaaccat	tgaaatgctg	900
aaggctcaga	attctgctgc	ccaggcgggt	attcagggag	cactgaatgg	tccagaccat	960
cctcccaaa	atcttcgcat	cagaagacag	cattcctctg	aaagtgttcc	tagtatcaac	1020
agtgccacaa	gccattccag	tattggcagt	ggtaattgat	ccgactccaa	gaagaagaaa	1080
aagaaaaact	gggtgaactc	tagaggaagt	gagctgagaa	gttctttcaa	acaagccttt	1140
gggaagaaaa	agtcacccaa	gcctccttca	tcacattctg	acattgaaga	gcttactgat	1200
tcattccctc	cggcatcccc	caagttaccc	cataatgctg	gtgactgtgg	ctcagcatcc	1260
atgaagccct	cacaatctgc	ttcagcgatc	tgtgaatgca	cagaagctga	ggcagagata	1320
attctgcagc	tgaagagcga	gctcagagaa	aaggaattaa	aattaacgga	tattcggctg	1380
gaggccctca	gctctgctca	tcattcttgat	cagatccggg	aagccatgaa	ccggatgcag	1440
aatgaaattg	aaatactgaa	agctgaaaat	gaccgggttg	aggcagaaac	tggttaacaca	1500
gctaagccta	ctcgccacc	gtcagaatcc	tcaagcagca	cctcctcttc	atcttccagg	1560
cagtcattag	gactttctct	aaacaatttg	aacatcacag	aggctgttag	ctcagatatt	1620
ttgctagatg	atgctggtga	tgcaactgga	cataaagatg	gccgcagtgt	gaaaattata	1680
gtctccataa	gcaagggcta	tggtcgagca	aaggaccaa	aatctcaggc	atatttgata	1740
ggatccattg	gtgttagtgg	aaaaaccaag	tggtggtgct	tagatggtgt	aataagacgt	1800
ctctttaagg	aatatgtatt	ccgaattgat	acatccacta	gccttggtct	gagctctgac	1860
tgcattgcta	gctactgtat	aggagactta	attagatccc	ataacctaga	agtgcctgaa	1920
ttgtgcctt	gtggatacct	tggtggagat	aataacatca	tcactgtgaa	cctcaaaggg	1980
gtagaagaaa	atagtttgga	cagttttgtt	tttgatacgc	tgattcctaa	accaattacc	2040
caaaggtact	ttactttgtt	gatggagcat	cacagaatta	tactctcagg	accgagtggt	2100
actggaaaga	cctatttggc	aaacaaactt	gctgaatatg	taataaccaa	atctggaagg	2160
aaaaaaacag	aggatgcaat	tgccactttt	aatgtggacc	acaagtcaag	taaggaattg	2220
caacaatata	tagctaacct	ggctgaacag	tgcatgtctg	ataataatgg	agtggagctc	2280
ccagttgtaa	taattcttga	taattctcat	catgtgggct	ctctgagtg	tatcttcaat	2340
ggttttctca	attgtaataa	caacaaatgt	ccatatatta	ttggaacaat	gaatcaggga	2400
gtttcttcat	caccaaactc	agagctgcac	cacaatttca	ggtgggtatt	atgtgcaaat	2460
catacagaac	cagtgaaaag	cttttttaggc	agatatcttc	gaagaaaact	catagagata	2520
gaaattgaaa	ggaacattcg	caataatgac	ctagtcaaaa	ttatagattg	gattccgaag	2580

14/56

Figure 1h (CONTINUED)

acgtggcatc	atctcaacag	ttttttggaa	acacacagtt	cttctgacgt	taccattggt	2640
ccccgactat	tccttccttg	ccccatggat	gtagaagggt	ctagagtatg	gttcatggat	2700
ctctggaaact	attcttttagt	accttatatt	ctggaggcag	tgagagaggg	tcttcagatg	2760
tatgggaaac	gcacaccatg	ggaagatcct	tcaaagtggg	tgcttgacac	atatccatgg	2820
agctcagcaa	ctctgcctca	ggagagccca	gccttacttc	agctgcgacc	agaagatggt	2880
gggtatgaaa	gctgcacatc	cactaaggaa	gccacaacct	caaagcacat	tccgcaaaact	2940
gacacagaag	gagatccccct	gatgaatatg	ctaataaagc	tccaagaagc	agccaattac	3000
tcgagcacac	aaagctgcga	cagcgaaagc	accagccacc	atgaagacat	tttgatttca	3060
tctcttgaat	ctaccctcta	gaggggtgaaa	aaagttaagg	gaaaagactt	tgctttttaa	3120
aaaatgtttc	aaaagaaagg	tatttttctac	aaaccactgc	cagtataaaa	gcaccctgtc	3180
aagggcccctg	accagaggtt	gtgggtctcca	aggaggcagc	agaactaagt	ctgaaccgcc	3240
aagatgctaa	attgcaatgg	aagcttaact	ttagtttatt	tctaagcatt	ttttatatct	3300
gtggagtaat	agaaagctcc	attactcaac	tggaaaggac	cctaatagaca	gggcaactga	3360
acagattgca	catgggatag	ccaaactgga	ctttctttgt	ttcctcttta	aaagtttaca	3420
atgcagacca	ttttttgtcc	cttcccttttg	tttccctctga	ggggctgttc	gccccaggca	3480
gggtccatct	ttctgatctg	tccaacctcc	tttgtgccac	acgggtgctgg	tcacagggct	3540
tcagtagtgt	ttgtgttgtg	cgctcacccc	attccagaac	aaatccaaga	ggccagtcct	3600
ccataagcac	aaatggaatt	gtgcaaccac	cagaaaaaca	ctactgtggc	aaactggaga	3660
agtgcgaatt	taattctaac	tgccacgttc	tcatgatgtg	ctccaccaac	tttttagtat	3720
atgagtcact	ggttttataa	ggttgttttt	accacagtgg	tcttttttaa	ccacctgccc	3780
actcccttaa	caagagtttt	ataccaatta	ttagtcaaca	ctgataaaag	gcttttttag	3840
ggctttatatt	gttttgagcct	tttcagtga	agaaggaaca	tttccatagg	tgctgtctca	3900
ctgccttaaa	acagatttct	atgacagttt	aacagttggt	ttaaatccta	aacctattgt	3960
aatttccact	gtcttttcat	ttacaaccaa	gcaacaccag	ttaacatagt	agcctctct	4020
ctatatatct	ttctcttttt	tttttttttt	tgaagaaatg	gataggagaa	agatcagtat	4080
tttttagcctt	gtgaatagat	cgctttgcct	atcctccaaa	atattaaaaa	aaccagaaa	4140
tgctctttga	ccgtcactta	aaacctaa	catgtggcga	aattccatcc	agttctaa	4200
gaaagagttt	cagaaggcag	gagattttga	attattatcc	agcagggctg	gaagcactag	4260
atgcagcatg	agcacaacta	ttcggtcttc	cttccctatt	gtttttgttt	ttttaatgag	4320
ttttgacgca	tggtgttttg	attgctattg	ttgtacatga	gaaattcagc	attaaagaac	4380
actgaagcgg	taaggtcact	gtggaagagg	aagcgtttat	actgtaaaag	aagggttagat	4440
ttgcacagtc	tactgggtag	gtattgtaaa	taataatttt	taaaacttgc	acaaatcaaa	4500
acaaacacaa	acaaaattgt	attttatcct	attggtgtta	agaggtgttt	cacttgctga	4560
gatttctctgt	acattgcaaa	caaatacaga	atgcaaaccc	tcaaagctgt	attatctggt	4620
gtgtttgtcc	tgtattttaca	gttgtttttg	actatgcagg	agctatcagt	gctagagtga	4680
gcatgcttca	aaactgtaca	tgaagccaat	atattttttg	ataagtaaaa	ctgtctgaaa	4740
gtacatctgt	catggcaggc	tttaaagaga	gtgcatgaaa	actgatcagt	cattggagaa	4800
gttaccacca	cacacaaagg	acaggtttta	agtttatgaa	acccaagggc	tagggcatgg	4860
tatagacttc	ttctatgagt	gtgtgaaaat	gtgttacttt	taggacgtgt	atttggtgct	4920
actctctgtg	accaccaatg	ggtcagttgc	tatagaacaa	caacaccacg	aaacatctgt	4980
gcagttttca	gagtgctaca	aagtcaatag	gtccttacac	ggtgctattg	cccaatggga	5040
aatccgaact	gaattttatgc	acatagaatt	gtcaccctga	ctttgaaagg	ccaaacatgg	5100
atcaaatctg	ttgtgaaaca	tcaatatatg	tagctggatg	agtgactagt	ttcccttgta	5160
taatatgtga	tctaagaaaa	ttgtctaatct	ttccctgcca	ttttgagaaa	cacagtccaa	5220
acatgagcat	aaacagaatt	tcctgcaata	catccagta	ggtccacctt	gtttacaact	5280
taaactagtt	tgtgaaacat	ttgtctgtat	acattttata	ttttgtacat	tttgatgtaa	5340
catatcatgt	aaataggcag	aaacagtga	ataaatcatc	tgaaaagttt	tgtagtcttt	5400
gtaaaagccc	aacaataagt	acttggtgtc	aatggactta	actggatgat	gtattttcta	5460
ttggttttatt	gttccctctag	cttgtaaac	agcttgcata	tatttttttg	caaagtgtga	5520
ccctgtatct	gtctaaatta	ttactttgcc	attaaagtgg	aattatttat	tgac	5574

Figure 1i. Overview of cloned nematode and human unc-53s variants



16/56

Figure 2: Illustration of a multiple sequence alignment between the different members of the Unc-53 protein family.

```

Ce-unc-53
Hs-unc-53/3      1 MPVLGVASKLRQPAVGSKPVHTALPIPNLGTTGSQKCSSRPLELAETESSMLSCQLALKS
Hs-unc-53/2
Hs-unc-53/1
                                     MTTSNVELIP.IYTDWANRHLSKGSLSKSIIRDISNDFRDYRLVSQLINV
Ce-unc-53
Hs-unc-53/3      61 TCEFGKKPLQCKAKEKEDSK.IYTDWANHYLAKSGHKRLIKDLQODIADGVLLAEITQI
Hs-unc-53/2      4 VSESSQQQKRKPVIHGLEDQKRIYTDWANHYLTKSGHKRLIKDLQODVTDGVLLAQIIQV
Hs-unc-53/1
Ce-unc-53      49 IVPINEFSPAFTKRLAKITSNLDGLETCLDYLNGLDCSKLTKTDIDSGNLGAVL
Hs-unc-53/3     120 IA..NEKVEDINGCPRSQSQMIENVVCLSFLLAARGVNVQGLSAAEIRNGNLKAIL
Hs-unc-53/2     64 VA..NEKIEDINGCPRNRSQMIENVIDACLNFLAAGKINIQGLSAAEIRNGNLKAIL
Hs-unc-53/1
Ce-unc-53     105 QLLFLLSTYKQKLRQLKKDQKLEQLPTSIMPPAVSKLPSPRVATSATASAT.....
Hs-unc-53/3     174 GLFFSLSKYKQ...QQHHQQQ..YQSLVLELQQRVTH.ASPPSEASQAKTQQQMQS(SLAA
Hs-unc-53/2     118 GLFFSLSKYKQ...QQQQPQK..QKLS.SPLPPAVSQVAGAPSQCQAGTPQQQVPV.TPQA
Hs-unc-53/1
Ce-unc-53     157 ..NPNSNFF.QMSTSRLOTPQ SRISKIDS..SKIGIKPKTSGLKPPSSSTTSSNNT.NSF
Hs-unc-53/3     226 ..RYATQSNHSGIATSQKKPT)RLPGP..SRVPAAGSSSKVQGA.....SNL..NRRSQSF
Hs-unc-53/2     172 PCQPHQAPAPHQSKAQAEQMS RLGGP.TARVSAAGSEAKTRGG.....STTANNRRSQSF
Hs-unc-53/1
Ce-unc-53     211 R.....PSSR.....SSGNMNVGSTISTSA.KSLESSSTYSSISNLR..
Hs-unc-53/3     277 NSIDKNK....PPNYANGNEKDS.SKGPOSSSG..VNGNVQPPSTAGQ.....PPAS
Hs-unc-53/2     226 NNYDKSKPVTSPPPPPSSHEKEPLASSASSHPG..MSDNAPASLESGSS.STPTNCSTSS
Hs-unc-53/1
Ce-unc-53     248 ..PT..SQLQKPSRPQTOLVRVATTTKIGSSK.....LAAPKAVSTPKLASVKTI.GAK
Hs-unc-53/3     322 AIPSP.SAS.KPWRSKSMNVKHSATSTM LTVKQSSTATSPTPSS...DRLKP.PVSEGVK
Hs-unc-53/2     283 AIPQPGAAT.KPWRSKSLSVKHSATVSM LSVK.....PPGPEA...PR....PTPEAMK
Hs-unc-53/1
Ce-unc-53     297 QEPDNSGGGGGGML.KLKLFPSSKNPSSSSNSP..QPT..RKAAAVP.....QQ.QTLSKI
Hs-unc-53/3     376 TAPSGQK....SMLEKPKLVNARTALRPPQPPSSGPGSDGGKDD..DAFESGEMEGFNSG
Hs-unc-53/2     329 PAPNNQK....SMLEKXLKLFNSKGGSKAGEGPGSRDTSCERLETLPSEEESELEAASRM
Hs-unc-53/1
Ce-unc-53     346 AAPVKSGLKPPPTS...KL...GSA.TSMSKLCTPKVS.....YRKTD.....
Hs-unc-53/3     430 LNSG..G...STNSSPKVSPKLAPPKAGSKNLSNKKSLLOPREKE.....EKNRDKNK.
Hs-unc-53/2     385 LTTV..G...PASSSPKIALKGIAQRTFSRALTNNKSSSLKGNKEKEKEKQREKDKESKD
Hs-unc-53/1
Ce-unc-53     381 .....
Hs-unc-53/3     478 .....VCTEK.PVREEKDQ.....VTEMAPKKTSKLASLIPKGSKTTAAKESLIP
Hs-unc-53/2     440 LAKRASVTERLDLKEEPKEDPSG...AAVPEM.PKKSSKIASFIPKGGKLNKAKKEPNAP
Hs-unc-53/1      1 ..MLPKRAKAPGGGGGMAFASAAELKVFKSGSVDSRVPGGPPASNLRKQKSLT
Ce-unc-53     381 .....APIISQQDSKRCSSKSSSEESGYAGFNSTSPSSSTEGSLM.HSTSSRSK
Hs-unc-53/3     523 SSSGIPKPGSKVPTVKQTIISPGSTASKESKFRITTKGSPSQSLSKP.IT.MEKASASSCP
Hs-unc-53/2     496 SHSGIPKPGMKSMPCSKSPAP..APSKEGERSRSGKLSSGLPQQKPLDG.RHSSSSSSS

```

17/56

Figure 2 (CONTINUED 1)

Ce-unc-53 429 TSDE.KSPSSDDLTNLASIVT.AIRQPIAATFVS.PNIIN.....KPVEE..KP.TLA
 Hs-unc-53/3 581 APLEGREAGQ..ASPSGS.CTMTVAQSSG..QSTGNG..AVQLP...Q.QQOHSHPNTAT
 Hs-unc-53/2 553 ASSEGKGPGG..TTLNHSISSQTVSGSVGTTQTGTGNTVSVQLP...QPQQQYNHPNTAT
 Hs-unc-53/1 106 FQAKGSPAGG..A....KTPLAPLAPNLGKPSRIIPRGPYAEVKPLSKAPEAAVSEDGKSD

 Ce-unc-53 476 VKGVKSTAKKDPFPAV..PFRDTQPTIG..V.VSPIMAHKKLTNDPVISEK....PEPE
 Hs-unc-53/3 630 VAPFIYRAHSENEGTLPSADSC.T.S.P..TKMDL..S.YSKTAKOCLEEISGE...GPE
 Hs-unc-53/2 608 VAPFLYRSQTDTECNV..TAESSS.T.G..VSVEP..SHFIKTGQPALEELTGE...DPE
 Hs-unc-53/1 160 DELLSSKAKAQKSSGPVPSAKGQE..E.RAFLKVDP...ELVVTVLGDLEQLLFSQMLDPE

 Ce-unc-53 526 ..KLQSMSIDTTDV.PPLP.PLKSVPPLKMTSIRQP.PTYD.....VLLKQKGI
 Hs-unc-53/3 680 TRRMRTVK.NIADLRQNLREETMSSLRGTOISHSTLE.TTFDSTVTTEVNGRTI.PNLTSRP
 Hs-unc-53/2 657 ARRLRTVK.NIADLRQNLREETMSSLRGTOVTHSTLE.TTFDTNVTTEMSSGRSILSLTGRP
 Hs-unc-53/1 215 SQKRRTVQ.NVLDLRQNLREETMSSLRGTOVTHSSLEMTCYDSD...DANPRSVSSLSNRS

 Ce-unc-53 570 T.....SPVKSPGY
 Hs-unc-53/3 738 TPMTWRLGQACPRLOAGDAPSLGAGY.PRSSTSRFIHTDPSRFMYTTPLRRAAVSRLGNM
 Hs-unc-53/2 715 TPLSWRLGQSSPRLQAGDAPSMGNGYPPRANASRFINTESGRYVYSAPLRRQLASRGSSV
 Hs-unc-53/1 271 SPLSWRYGQSSPRLQAGDAPSVGGSCRSEGTPAWYMHGERAHYSHTMPMR..SPSKLSHI

 Ce-unc-53 579 EQSSASEDSIVAHASQVTPPTKTSNGHSLERRMGKN.KTSE..SSGYTS DAGVAMCAKM
 Hs-unc-53/3 797 SQIDMSEKA..SS.DLDMSS.EVDVGGYMSDGDILGKSLRTDD..INSGYMTDGGNLNLYTRS
 Hs-unc-53/2 775 CHVDVSDKA..GD.EMDLEGISMDAPGYMSDGDVLSKNIRTD..ITSGYMTDGGGLGLYTRR
 Hs-unc-53/1 329 SRLELVESLSD.EVDLKS.....GYMSDSDLMGKTMTEDDDITG.....

 Ce-unc-53 636 REKLKEYDDM..TRRA..QN....GYPDNFEDSSSLSSGISDNNELDDISTDDLSGV..
 Hs-unc-53/3 853 LNRIP..D.TATSRDIIQRGVHDVTVADADSWDDSSSVSSGLSDT..LDNISTDDLNNTSS
 Hs-unc-53/2 832 LNRLP..DGMVAVRETLQRNTSLGLGDADSWDDSSSVSSGISDT..IDNLSDDINTSSS
 Hs-unc-53/1 369WDESSSISSGLSDA..SDNLSSEEFNASSS

 Ce-unc-53 685D..MATVASKHS.....
 Hs-unc-53/3 908 VSSYSNITVPSRKNTQ..LATDSEKRSTTDET..WDSPEELKKPEE.DPDS...HGDAG..
 Hs-unc-53/2 888 ISSYANTPASSRKNLD..VQDAAEKHSQVERNLSW.SGDDVKKSDG.GSDSG.IKMEPG..
 Hs-unc-53/1 397 LNSLPSTPTASRRNSTIVLRDSEKRS LAESGLSWFSESEEKAPFKLEYDSGSLKMEPGT

 Ce-unc-53 695
 Hs-unc-53/3 959 GKWKTVSSGLFEDPEKA.GQKASLSVSQTGSWRRGMSAQGG..AP.SRQKAGTSALKTP..
 Hs-unc-53/2 942 SKWRRNPDSVSDSDKSTSGKKNPVISQTGSWRRGMTAQVGITMPRTKASAPAGALKTPG
 Hs-unc-53/2 E
 Hs-unc-53/1 457 SKWRRERPESCDSSSKGGELKKPISLGHPGSLKKGKTPPVAVTSPITH..TAQSALKV..

 Ce-unc-53 695
 Hs-unc-53/3 1014 .GKTDDAKASEKGPAPLKCSSLQSPSDAGKSSGDEGK.PSGIGRSTA.TSSFGFKKP
 Hs-unc-53/2 1002 TGKTDDAKVSEKGRSPKASQVKRSPSDAGRSSGDESKKPLPSSSRTPANANSFGFKKQ
 Hs-unc-53/1 513 AGK.PEGKATDKGKLAVKNTGLQRSSSDAGRDRLSDAK.PSGIARP.STSGSFCYKRP

 Ce-unc-53 695
 Hs-unc-53/3 1071 SGVGSS.AMITSSGATITSGSATLGKIPKSAAGGKSNAGRKTSLDGSONQDDVVLHVSS
 Hs-unc-53/2 1062 SGSAAGLAMITASGVTVTSRSATLGKIPKSSALVRS.AGRKSSMDGAQNQDDGYLALSS
 Hs-unc-53/1 570 P.PATGTATVMQTG.....GSATLSKIQKSSGIPVKPVNGRKTSLDVNSAEPGFLAPGA

 Ce-unc-53 695
 Hs-unc-53/3 1130 KTTLQYRSLPRPSKSSSTSGIPGR.GGHRSSSTESID.SNVSSKSAGATTSKLREPTKIGSG
 Hs-unc-53/2 1121 RTNLQYRSLPRPSKSNR...NG.AGNRSSTSSID.SNISSKSAGLPVPKLRPSKTALG
 Hs-unc-53/1 624 RSNITQYRSLPRPAKSSSMVSTGGRGGPRPVSSSIDPSLLSTKQCGLTPSRLKEPTKVASC

18/56

Figure 2 (CONTINUED 2)

Ce-unc-53 695 VAVSDSESVLSGSPKSSPTSASACG.AQGLRQPGS
 Hs-unc-53/3 1188 .RSS.FVIVNQTDKEKEK.....VAVSDSESVLSGSPKSSPTSASACG.AQGLRQPGS
 Hs-unc-53/2 1176 .SSL.PGLVNQTDKEKG.....ISSDNESVASCNVKNPAAQPVSSPAQTSLOPGA
 Hs-unc-53/1 684 .RTT.PAPVNQTDREKEKAKAXAVALDSDNISLKSIGSPESTPKN.....QASHPTAT

Ce-unc-53 695 DYSHFVRHPTSSSSSKPRVP
 Hs-unc-53/3 1239 KYPDIASPTFRR(LFGAKAGGKSASAPNTEGVKSSVMPSPSTTLARQGSLESPPSSGTGSM
 Hs-unc-53/2 1226 KYPDVASPTLRR LFGGKP.TKQVPIATAENMKNSVVISNPHATMTQOQNLDSP.SGSGVL
 Hs-unc-53/1 735 KLAELPPTPLRA T.AKSFVKPPSLANLDKVNNSN.....SLDLPS

Ce-unc-53 714 SRSSTSVDSKSRAEQENVYKLLSOCRTSQRGAAATSTFGQHSLSRSPG.....YSSYS
 Hs-unc-53/3 1299 GSAGGLSGSSSPLFNKPSDLTTDVISLSHSLAS....SPASVHSFTSGGLVWAANMSSSS
 Hs-unc-53/2 1284 S....SGSSSPLYSKNVDLN.....QSPLAS....SPSSAHSAPSNSLTWGTNASSSS
 Hs-unc-53/1 774SDTHASKVPDLHATSSASGGFLPSCITPSPAPILNINSASFSGGLELMGPF

Ce-unc-53 766 FHLVSADKDTMS.MHSQTSRRPSSQKPSYSG(QFHSLSDRKCHLOEF.TSTEHMAALLSP
 Hs-unc-53/3 1355 AGSKDTPSYQSMSTSLHTSSESIDLPLS....HMGSLSGLTGTGTHEVQSL..LMR.TGVS
 Hs-unc-53/2 1329 AVSKDGLGQSVSSLSHTSCESIDISLSSGGVP SHNSSTGLIASSKD.DSLTPFVR.TNSV
 Hs-unc-53/1 826 SVPKETRMYPKLSGLHRSMESLQMPMS.....LPSAFPSTFPVPTPP.APPA

Ce-unc-53 824 RRVPNMS KYDSS)(AAALNASGMSRSMILLES L SFRPFRHQSPADS CIITASPSAPRRS
 Hs-unc-53/3 1407 RSTL.SES).....(MQLDRNTLPKKGLR)YTPSSRQANQEEG
 Hs-unc-53/2 1387 KTL.SES .PLSS PAASPKFCRSTLPRKQSD PHLDRNTLPKKGLR YTPTSQRLTQEDA
 Hs-unc-53/1 872 APTE.EET .EELT WSGSPR.....AGQLDS NQDRNTLPKKGLR Y....QLOSQEET

Ce-unc-53 883 HSPRGPTARIPLSL.ASSPVHVNNW)GSYSARSQGSST CIYGETF.....
 Hs-unc-53/3 1441 KEWLRSHSTGGLQDTGNQSPLVSPSAM SSSAAGKYHFSNL VSPTNLSQFNLPGPSMMRSN
 Hs-unc-53/2 1444 KEWLRSHSAGGLQDTAANSFFSSGSSV TSPSGTRFNFSQL ASPTTVTQMSLSNPTMLRTH
 Hs-unc-53/1 918 KERPHSHTIGGLPESDDQSELPSPAL PMSLSAKQQLTNI(VSPTAAT....TPRITRSN

Ce-unc-53 928OLHRLS...DEKSPAHSKSEM.....GSQLSLASTT..AY
 Hs-unc-53/3 1501 SIFAQDESFDLYDDSQLCGSATSLEERPRAS.HSGSFRDSMEE VHGSLSLVSSTSSLY
 Hs-unc-53/2 1504 SLNADGQYDPTDSRFRNSSMSLDEKSRMS.RSGSFRDGFEE VHGSLSLVSSTSSVY
 Hs-unc-53/1 973 SIPTHEAAFELYSGSQM.GSTLSLAERPCKMI.RSGSFRDPTDD)VHGSVLSLASSASSTY

Ce-unc-53 959 GS LNEKYEHA .IRDMARDLECYKNTVDSLTKKQ.....
 Hs-unc-53/3 1560 ST AEEKAHSE QIHKLRRELVASQEKVATLT.SQLSAN
 Hs-unc-53/2 1563 ST PEEKCQSE .IRKLRLRELDASQEKVSALT.TQLTAN
 Hs-unc-53/1 1031 SS(AEERMQSE)QIRKLRLRELESSQEKVATLT.SQLSAN(VSAMKYGKIKAVLITIVRQVQPR

Ce-unc-53 991 .ENYG A.LFDLFEOKLRKLTQHIDRSNLKPEEAI RFRQDIAHLRDISNHLASNSAHANEGAG
 Hs-unc-53/3 1596AHLVAAFEKSLGNMTGRLOSLTMTAECK...ESELIELRETIEMLKAQNSAAQAAIQ
 Hs-unc-53/2 1598AHLVAAFEQSLGNMTIRLOSLTMTAECK...DSELNELRKTIELLKKQNAQAQAIN
 Hs-unc-53/1 1090 EENYL)ANLVAAFEQSLVNMTSRLRHLAETAEEK...DTELLDLRETIDFLKKKNSAQAVIQ

Ce-unc-53 1041 ELL.....RQPSLESVASHRSSMSSSSKSSKQEKISLSSFGKNR
 Hs-unc-53/3 1650 GALNGPDHPPK.....DLRIIRQHSSESVSINSATSHSSIGS...GNDADSKKKKK
 Hs-unc-53/2 1652 GVINTPELNCKGNGTAQSADLRIRQHSSESVSINSATSHSSVGS...NIESDSKKKKR
 Hs-unc-53/1 1149 GALNASETTPK.....ELRIIRQNSSEDSSISLNSITSHSSIGS...SKDADAKKKKK

Ce-unc-53 1090 KSW(ALSVDSQ)IRSSLK.FTKKKN.K....NYD.....EAHMPs...I...S.GSQG..
 Hs-unc-53/3 1699 KKW(VNSRGSE)LRSSFQAFGKKKSTKPPSSHSIDIEELT...DSSLPASPKLPHNAGDCGSA
 Hs-unc-53/2 1709 KKW(VN....E)LRSEFQAFGKKKSPKSASSHSIDIEEMT...DSSLPSSPKLPHN.GSTGS..
 Hs-unc-53/1 1198 KSW(V....YE)LRSSFNKAFSIKKCPKSA.SSYSDIEEIATPDSSAPSSPKLQHGSTETASP

Ce-unc-53 1129T.L.....DN.....ID....VIELKQELKERDSALY

19/56

Figure 2 (CONTINUED 3)

Ce-unc-53 1151 EVRLDNLDRAREVDVLRSTVNRLKTENKQLKKEVDKL...TNGP...ATRASSRAS...I.
 Hs-unc-53/3 1816 DIRLEALSSAHHLDOIREAMNRMQNEIEILKAENDRLKAETGNTAKPTRPPSESSSSSTSS
 Hs-unc-53/2 1800 DIRLEALSSAHQLDQLREAMNRMQSEIEKLKAENDRLKSESQSGS.CSRAPSQVS...IS
 Hs-unc-53/1 1310 DIRLEALNSAHQLDQLRETMHNMQLLEVDDLKAENDRLKVAPGPSSSGST..PGQVPGSSAL

 Ce-unc-53 1202 ...PVIYD...DEHVDYDAACSSTS.ASQSSKRSSSGCNSIKVTNV..DIAGEI SS
 Hs-unc-53/3 1876 SSSR.QSLGLSLNNLN.ITZAVSS DILLDDAGDATGHKD.GRSVKIIVSISKGYGRAK DQ
 Hs-unc-53/2 1856 ASPR.QSMGLSQHSLN.LTESTSL(DMLLDDTGEC SARKEGGRHVIVVSFQEMKWKE)DS
 Hs-unc-53/1 1358 SSPR.RSLGLALTHSF.GPSLADT DLSPMDGISTCGPKEE.VTLRVVVRMPQHITKG DL

 Ce-unc-53 1248 IVNPDKEIIVGYLAMS TSQSCWKDI.DVSILGLFEVYLSRIDVEHOLGIDARDSILGYOI
 Hs-unc-53/3 1933 ..KSQA.YLIGSIGVS .GKTKW.DVLDGVIRRLFKEYVFRIDTSTSLGLSS.DCIASYCI
 Hs-unc-53/2 1914 ..RPHL.FLIGCIGVS(*).GKTKW.DVLDGVVIRRLFKEYIIHVDPVSQGLNS.DSVLGYSI
 Hs-unc-53/1 1425 ..KQQE.FFLGCSKVS .GKVZW.KMLDEAVFQVFKDYISKMDPASTLGLST.ESIHGYSI

 Ce-unc-53 1307 GELRAVIGDSTTMITSH..PTDILT.SSTTIRMFHGAQAQSRVDSVLDMLLPKOMILQL
 Hs-unc-53/3 1987 GDLIR....SHNLEVPELLPCGYLVGDNNIITVNLKGVEENSLDSFVFDTLIPKPITORY
 Hs-unc-53/2 1968 GEIKR....SNTSETPELLPCGYLVGENTTISVTVKGLAENSLDSLVFESLIPKPILORY
 Hs-unc-53/1 1479 SHVKR....VLDAEPPPEPPCRR..GVNN.ISVSLKGLKEKCVDSLVFETLIPKPMQHY

 Ce-unc-53 1364 VKSILTERRVLVLAGATGIGKSKLAKTLAAYVSIRTNQS.EDSIV.NISIPENNKEELLOQ
 Hs-unc-53/3 2042 FNLLMEHHPILSGPSGTGKTYLANKLAEYVITKSGRKTEDAIATFNVDHKSSKELQO
 Hs-unc-53/2 2023 VSLLIEHRRIILSGPSGTGKTYLANRLSEYIVLREGRELTGVIATFNVDHKSSKELRQ
 Hs-unc-53/1 1531 ISLLLKHRRVLVLSGPGSGTGKTYLTNRLAEYLVERSGREVTEGIVSTFMHQQSCKDLOL

 Ce-unc-53 1421 VERRLEKIDRSKESC....IVILDNIPKNRIAFVVSVFANVPLQN...NEGPFVVCVTN
 Hs-unc-53/3 2102 YLANLAEQCSADNNGVELPVVILDLN..HHVGSLSDF.NGFL.NCKYKCPYIIGTMN
 Hs-unc-53/2 2083 YLSNLADQCENSENNAVDMLVILDLN..HHVSSLGEIF.NGLL.NCKYKCPYIIGTMN
 Hs-unc-53/1 1591 YLSNLANQIDRETGIGDVPLVILLDDL..SEAGSISELV.NGAL.TCKYKCPYIIGTMN

 Ce-unc-53 1473 R..YQIPELQIHNFMSVMSNRLE..GFILRYLARRAVEDEYRLTVQMPSELFKII
 Hs-unc-53/3 2158 QGVSSSPNLELHNFWRVLCANHTFVKGFLGRYLARKLIEISIERNIRNN.DLVKII
 Hs-unc-53/2 2139 QATSSTPNLQLHNFWRVLCANHTFVKGFLGRFLRRKLMETEISGRVRNM.ELVKII
 Hs-unc-53/1 1647 QPVKMTPNHGLHLSFRMLTFSNVVEPANGFLVRYLRRKLVESDSDINANKE.ELLRLV

 Ce-unc-53 1526 DFFPIALQAVNNFIEKTNSVDVTVGPRACLNCPITVDGSKWFIRLWNENFIPLYLRAV
 Hs-unc-53/3 2215 DWIPKTHHLNLSFLETHSSSDVTIGPRFLPCPMDEGSRVWFMDLWNYSILVPILEAV
 Hs-unc-53/2 2196 DWIPKTHHLNLSFLEAHSSSDVTIGPRFLSCPIDVDGSRVWFTDLWNYSIIPYLLEAV
 Hs-unc-53/1 1704 DWVPKLWYHLHTFLEKHSTSDFLIGPCFFLSCPIGIEDFRTWFDLWNYSIIPYLQEGA

 Ce-unc-53 1585 RDGKXTFGRCTSFEDFTDIVSKKWPWFGENPEN...VLXRLQLQDL.....VPSPAN
 Hs-unc-53/3 2274 REGLQMYGKRTFWEDPSKWVLDITYFW..SSATLPQESFALLQLRPEDVGYESCTSTKEAT
 Hs-unc-53/2 2255 REGLQLYGRAPWEDPAKWMDITYPW..AASPQHEWPLLQLRPEDVGFDGYSMPREGS
 Hs-unc-53/1 1763 KDGIVHGXKAAWEDPVEWVRDTLFW..PSA..QQDQSKLYHLEPPTVGPHSIASPPEDR

 Ce-unc-53 1635 SSRQ.....HFNPL.ESLIQL.HATKH...QTIDNI
 Hs-unc-53/3 2332 TSKHIPQTDTEGDPLMMLMLKQEAANYSSQSCDSE..STSHHEDILDSSLESTL
 Hs-unc-53/2 2313 TSKQNPSSDAEGDPLMMLMLKQEAANYSSQSYDSDSNSNSHHDDILDSSLESTL
 Hs-unc-53/1 1819 TVKDSTPSSLDSDPLMMLMLKQEAANY..IESPDRET.....ILDPNLQATL

20/56

Figure 3: Illustration of a multiple sequence alignment between *C. elegans* Unc-53 (Ce) and *C. Briggasae* Unc-53 (Cb).

```

Cb 1 MTTSNVELIPIYTDWANRHLKSGALSRIIDISNEFRDYRLVSQLINVIVFINEYSPTYTYPLAKITSNLDGLETCLOYL
Ce 1 MTTSNVELIPIYTDWANRHLKSGSLEKSIIRDISNDFRDYRLVSQLINVIVPIKEFSPAFTKRLAKITSNLDGLETCLOYL

Cb 81 KNLGLDCSKLTKTDIDSGNLGAVLQLLFLLSYKQKRLRLKQKLEQLFVTTTTTAIMPFAVSNIPYSRLPSRPVPPA
Ce 81 KNLGLDCSKLTKTDIDSGNLGAVLQLLFLLSYKQKRLRLKQKLEQLF.....TSIMPFAVSKLFSRPVATSATASA

Cb161 SNFNSNFTQMSTSRLOTQSRISKFDSTKIGIKPKTTSGLRPP. STTSNTNINSFRPESFSSGNVNVGSTISTSARSLD
Ce156 TNFNSNFTQMSTSRLOTQSRISKIDSSKIGIKPK. TSGLKPPSSSTTSNTNNSFRPESRSGNVNVGSTISTSAKSLE

Cb240 SSSAYSSISNLSRFTPSSTQIQKPTSRLOTQOVVATTYKIGSSKLAAPRAVSTPKLASVKTIKTTTEHDNS....GGML
Ce235 SSSYSSISNLSNRPT...SQLQKP. SRPQTQLVRVATTYKIGSSKLAAPRAVSTPKLASVKTIGARQEPDSSGGGGGGMML

Cb316 KKLKLFSSKNASSSNNEPQFLRKA....EQ...SKIAAPVKTGLKPPTSTTNKLGSLTSMKSLCTPKVSYRKPDILLHTKS
Ce311 KKLKLFSSKNPSSSSNSPQPTRKAAAVPQQOTLSKIAAPVKSGLKPPTS...KLGSLTSMKSLCTPKVSYRKPDAPITISQQ

Cb389 DSKRCSKSSSEESGYAGFNSTSPASSSTEGSLSMHSTSSKSTSDSEKPSDDLTNLSIVTAIRQPIATSAVSP.VISK
Ce388 DSKRCSKSSSEESGYAGFNSTSPSSSTEGSLSMHSTSSKSTSDSEKPSDDLTNLSIVTAIRQPIATPVSPNIIK

Cb468 PVEEKPTLAVKGV.SASEKLPPTVTERTNQPTIGVVSFIMAEKLFSESTPSEKVDNPEKISSMSID.CDLPPPPTVL
Ce468 PVEEKPTLAVKGVKSTAKDPPFAVFPRTQPTIGVVSFIMAKRLTNDPVISEX...PEPEKLSMSIDTDDVPPLE.PL

Cb546 KSLERVPPKMTPIRQPTTYDVLVKLGKITSPVKSFGYDQVESASSEDSIVAH...VQMAPPVQKTSAGQSSMERRIQKKT
Ce545 KSV...VPLRMTPIRQPTTYDVLVKLGKITSPVKSFGYEQ...SSASEDSIVAHASQVTPPT.KTS.GNHSLERRMGKNT

Cb624 SESSGYASDAGVAMCAKREKLKEVDMTRRAQNGYPDNFEDSSSLSGISDMNELLDDISTDDLSGIDMATVASKHSDYS
Ce619 SESSGYASDAGVAMCAKREKLKEVDMTRRAQNGYPDNFEDSSSLSGISDMNELLDDISTDDLSGIDMATVASKHSDYS

Cb704 HFVRHTSSSSSRPRVPSRSTSVDSRSRVEQENVYKLLSQCRTSQRGAATATSSFGQHSLSRSPGYSSYPHLSVADKDT
Ce699 HFVRHTSSSSSRPRVPSRSTSVDSRSRVEQENVYKLLSQCRTSQRGAATATSSFGQHSLSRSPGYSSYPHLSVADKDT

Cb784 MSMHSQTSRRPSSQKPSYAGQFHSLEKCHLQEFSTAHRMAALLSPRVPNSMSKIDSSSGYSARSRCGSSTGIYGEF
Ce778 MSMHSQTSRRPSSQKPSYAGQFHSLEKCHLQEFSTAHRMAALLSPRVPNSMSKIDSSSGYSARSRCGSSTGIYGEF

Cb864 FQLHRLSDEKSPAHSAKSEMSQLELASTTAYGSLNEXYEHATRMARDLECYKNTVLSLTKKQENYGFALFDFEQKLRK
Ce857 FQLHRLSDEKSPAHSAKSEMSQLELASTTAYGSLNEXYEHATRMARDLECYKNTVLSLTKKQENYGFALFDFEQKLRK

Cb944 LTHSIDRSNLEKFEATFRQDIAMLEIINHLATNSMIVNEGAGELLRQPSLESVASHRSSMSSSSKSKQEKISLSSFG
Ce937 LTHSIDRSNLEKFEATFRQDIAMLEIINHLATNSMIVNEGAGELLRQPSLESVASHRSSMSSSSKSKQEKISLSSFG

Cb1024 KMKKSWIRSLSKFTKKKNKNYDEGHMPSISGSQGTLDNIDVIELKQELKERDSALYEVRLDNLDRAREVTVLKETVNLK
Ce1017 KMKKSWIRSLSKFTKKKNKNYDEGHMPSISGSQGTLDNIDVIELKQELKERDSALYEVRLDNLDRAREVTVLKETVNLK

Ce1104 KLENKQLKKEVDKLTNTSTTRASSRSLFYIQDDEHVYDHACSSTSSASQSSKRSQCNISIKVTNVNDIAGEISSIVNPDK
Ce1097 KLENKQLKKEVDKLTNTNGPATRASERASIPVYDDEHVYDAACSSSTSSASQSSKRSQCNISIKVTNVNDIAGEISSIVNPDK

Cb1184 EIIIVGYLEMPANSTWKDIEDSTILDSPEKYLKIDLDRLGLDAKDAIFGYQIGELRRVIGDESTIITSHPVDILTPTTT
Ce1177 EIIIVGYLAMSTSQSCWKDIDVSIQLFEVYLSRIDVEHQICIDARDSILGYQIGELRRVIGDESTIITSHPTDILTSTT

Cb1264 IRMFMYGAAQSRVDSMVLDMLLPRQMLQLVKSITERRRLVLGATGIGKSKLAKTLAAVYSLQTNQSEDKIVNITIPEN
Ce1257 IRMFMYGAAQSRVDSMVLDMLLPRQMLQLVKSITERRRLVLGATGIGKSKLAKTLAAVYSLQTNQSEDKIVNITIPEN

Cb1344 NKEELLQVERRLKILRSKEACVITLDNIPKNRIAFVVSFANVPLQNNEGPFVVTNRYQIPELKTIIPFKMSVMSNR
Ce1337 NKEELLQVERRLKILRSKESCTVITLDNIPKNRIAFVVSFANVPLQNNEGPFVVTNRYQIPELQIHNFKMSVMSNR

Cb1424 LEGFILRYLRRRAVEDEYRLSVQMPSELRIIEFFPVALQAVNNFIEKTNVDVTUGPRACLNCFITDGSREWFIRLWN
Ce1417 LEGFILRYLRRRAVEDEYRLTVQMPSELFIIDFFPIALQAVNNFIEKTNVDVTUGPRACLNCFITDGSREWFIRLWN

Cb1504 QNFIPYMERVARDCGKTLGRCTSFEDPTDIVSKWPFDCPNFEDVLKRLQLQDLAPSPANSRQPFNPLESILQLHATK
Ce1497 QNFIPYMERVARDCGKTLGRCTSFEDPTDIVSKWPFDCPNFEDVLKRLQLQDLVSPANSRQHFNPLESILQLHATK

```

21/56

Figure 4. Prosite Signatures

Block A. Large family:

IYTDWANXXLX(K,R)(A,G,S,T)XXX(K,R)X(ILVA)(H,K,R,T,S)D(I,L)XXDXDXXL(L,V)
)(A,S)(N,D,Q,E)(I,L,V,A)I(N,D,Q,E)(I,L,V,A)(I,L,V,A)(V,A,T,S)X(17,19)(
 I,L,F)(N,D,Q,E)X(I,L,V,A)(N,D,Q,E)XCLXXLXXX(A,G,S,T)(I,L,V,A)X(4,5)(I,
 L,V,A)(S,T)XX(N,D,Q,E)IXXGXLXA(V,I)LXL(L,F)FXLSX(Y,F)KQ

Block B. Vertebrate:

PEXXRXRTV(Q,K)N(I,L,V,A)(I,L,V,A)DLRQNLLEETHSSSLRG(S,T)Q(V,I)(S,T)HS(S,T)
)LEX(0,1)T

Block C. Vertebrate:

RX(S,T)P(L,M)(S,T)WRXGQ(S,A)XPRLOAGDAPS

Block D. Vertebrate:

GYMSDXD(M,L,V,I)(M,L,V,I)(A,G,S,T)KXXXD(2,3)I(N,T)(A,G,S,T)G(Y,-)

Block E. Vertebrate:

WD(D,E)SSS(M,L,V,I)SSG(L,I)SDXXDN(L,I)S(S,T)(D,E)(D,E)XN(A,G,S,T)(S,T)
 SS

Block F. Vertebrate:

DRNTLPKXGLRY

Block G. Large family:

GSX(I,L,V,A)SL(I,L,V,A)S(A,G,S,T)(A,G,S,T)S(0,2)XY(A,G,S,T)XX(E,N)E(K,
 R)X(4,5)I(R,H)X(L,M)XR(D,E)LXXXXXXVXXLTXXXXXXXLXXXFE(Q,K)(S,K)LXXXTXX
 (L,I)XX(L,S)XXXXE(Q,E)X(3,6)(D,E)(L,I)XXLRXXX(N,D,Q,E)XLXXXX(A,S)XA(N,
 D,Q,E)XXXXXX(L,I)X(0,21)RQXSX(N,D,Q,E)S(I,V)XSXXSXXSXXSX(A,G,S,T)S

Block G. Vertebrate:

SGSFRDXX(D,E)(E,D)VHGSXLSL(V,A)SS(T,A)SSXYS(T,S)XEE(K,R)XXSE(Q,-
)I(R,H)KLRRELX(A,S)SQEKVX(T,A)LT(T,S)QL(S,T)ANAXLVAAFE(Q,K)SLXN(M,L,V,
 I)MTXRL(Q,R)XLXXTAE(Q,E)KXXELXXLRXTI(D,E)XLKXXN(A,S)XAQAXIXGX(L,I)N(A,
 G,S,T)X(N,D,Q,E)XXXXX(C,8)(N,D,Q,E)LRI(K,R)RQXSS(N,D,Q,E)S(I,V)SS(I,L)
 NSXTSHSSXGS

Block H. Large family:

(V,L)DSXVX(D,E)XL(I,L)PKX(M,L,V,I)XXXXXXX(L,I)(M,L,V,I)XXXR(I,L)(I,V)L
 (A,S)GX(T,S)GXGK(T,S)XL(A,T)XXLXXY(M,L,V,I)XX(R,K)

and

P(E,N)XX(I,L)HXXF(K,R)XXX(A,S)NXXEX(0,3)GF(L,I)XR(Y,F)L(K,R)(K,R)(K,R)
 X(M,L,V,I)(D,E)

and

F(I,L)EXXX(T,S)X(D,E)XXXGPKXX(L,I)XCP(M,L,V,I)X(V,I)(D,E)XX(R,K)XWFXXL
 WNXXX(I,V)PY(L,I)XXX(A,V)(R,K)(D,E)GXXXXGXX(T,A)X(F,Y,W)EDP

Block H. Vertebrate:

(V,L)DSXVF(D,E)(T,S)DIPKP(M,L,V,I)XQYXXLL(M,L,V,I)XHXR(I,L)(I,V)LSGPS
 GTGKTYL(A,T)NRLXEY(M,L,V,I)XX(R,K)GR

and

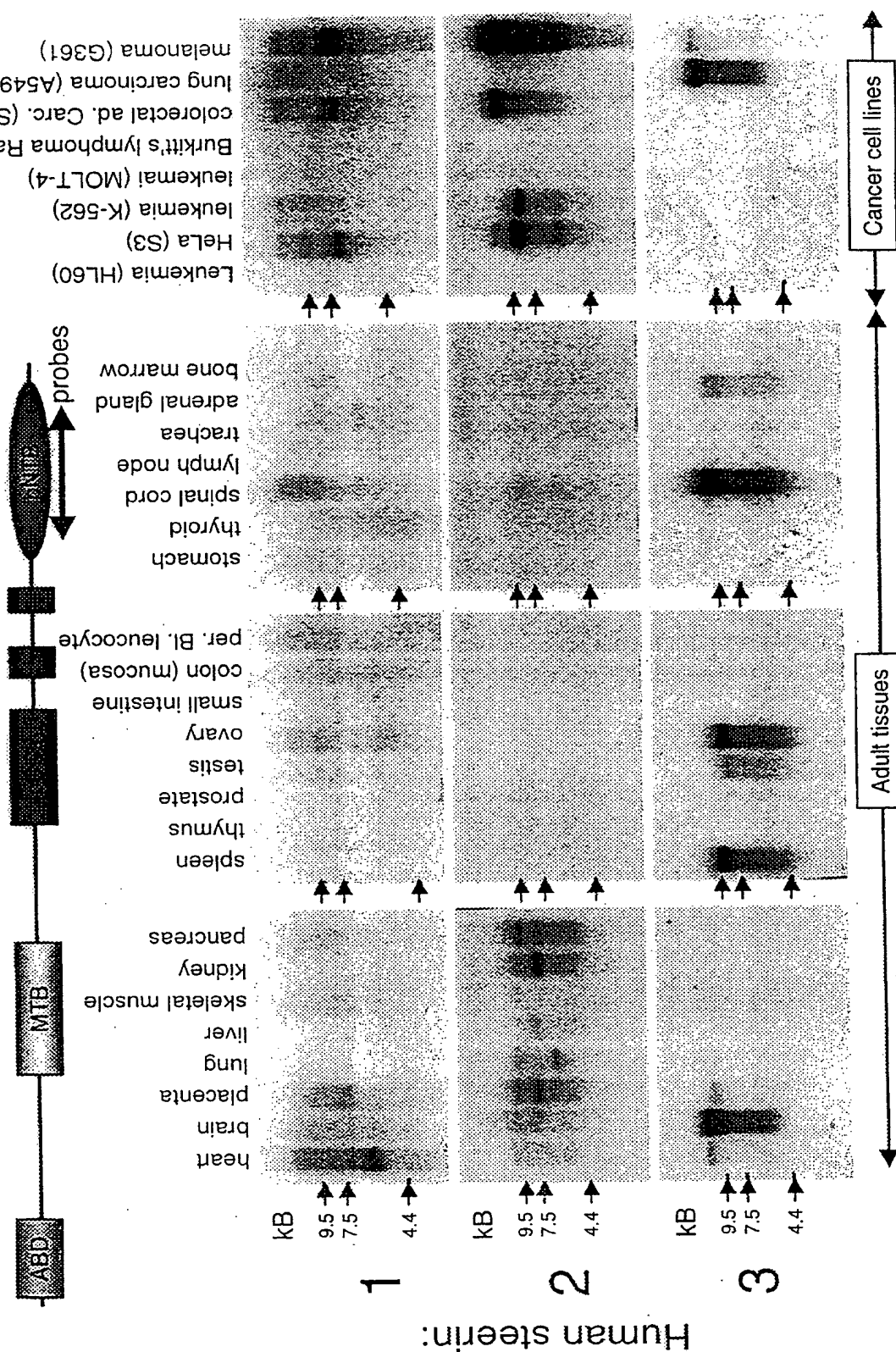
VI(I,L)LD(D,N)LXXXXS(I,L)XX(I,L)XNGXLXCKYXKCPYIIGT(T,M)NQXXXX(T,S)PNXX
 LHXXFRXXXX(A,S)NXXEP(A,V)XGFLXR(Y,F)L(K,R)(K,R)(K,R)L(M,L,V,I)(D,E)

and

(R,K)(V,I)(L,I)DWXPKXWXH(I,L)XXFLEXHS(T,S)SDXXICPXXFLXCP(M,L,V,I)X(V,I)

22/56

FIG. 5a Expression of Hs-unc-53 in tissues and cancer cells by Northern blotting



23/56

FIG. 5b Differential expression of *Hs-unc-53/3* in human brain regions

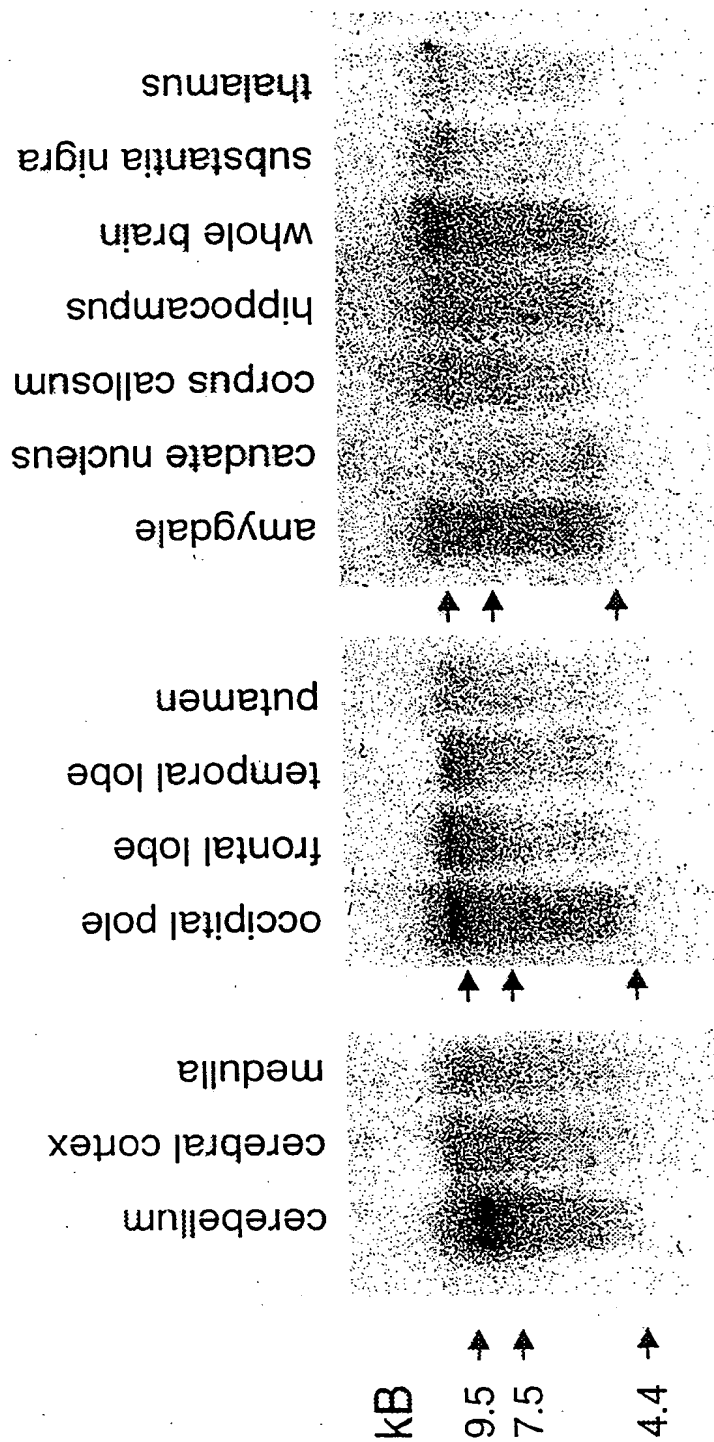
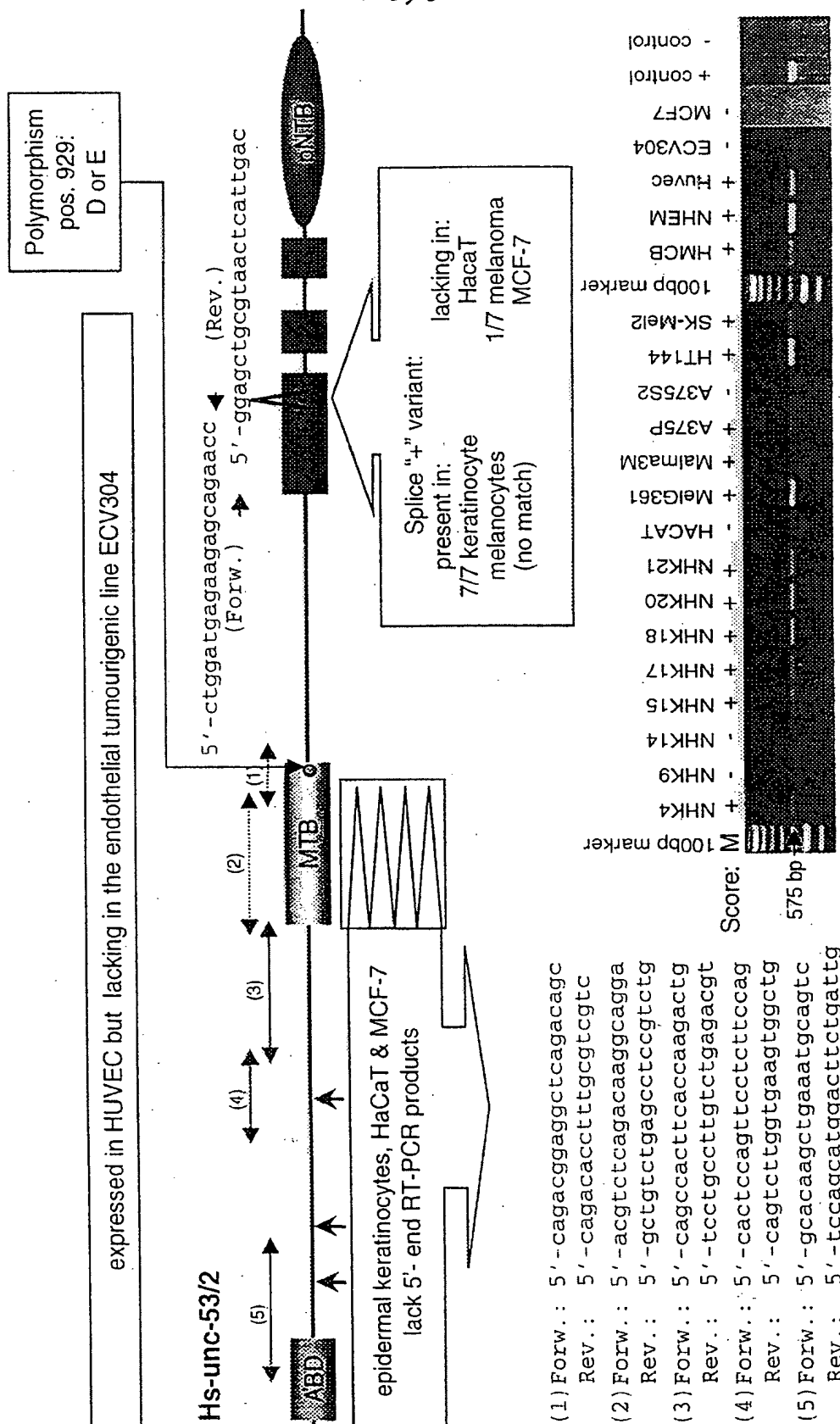


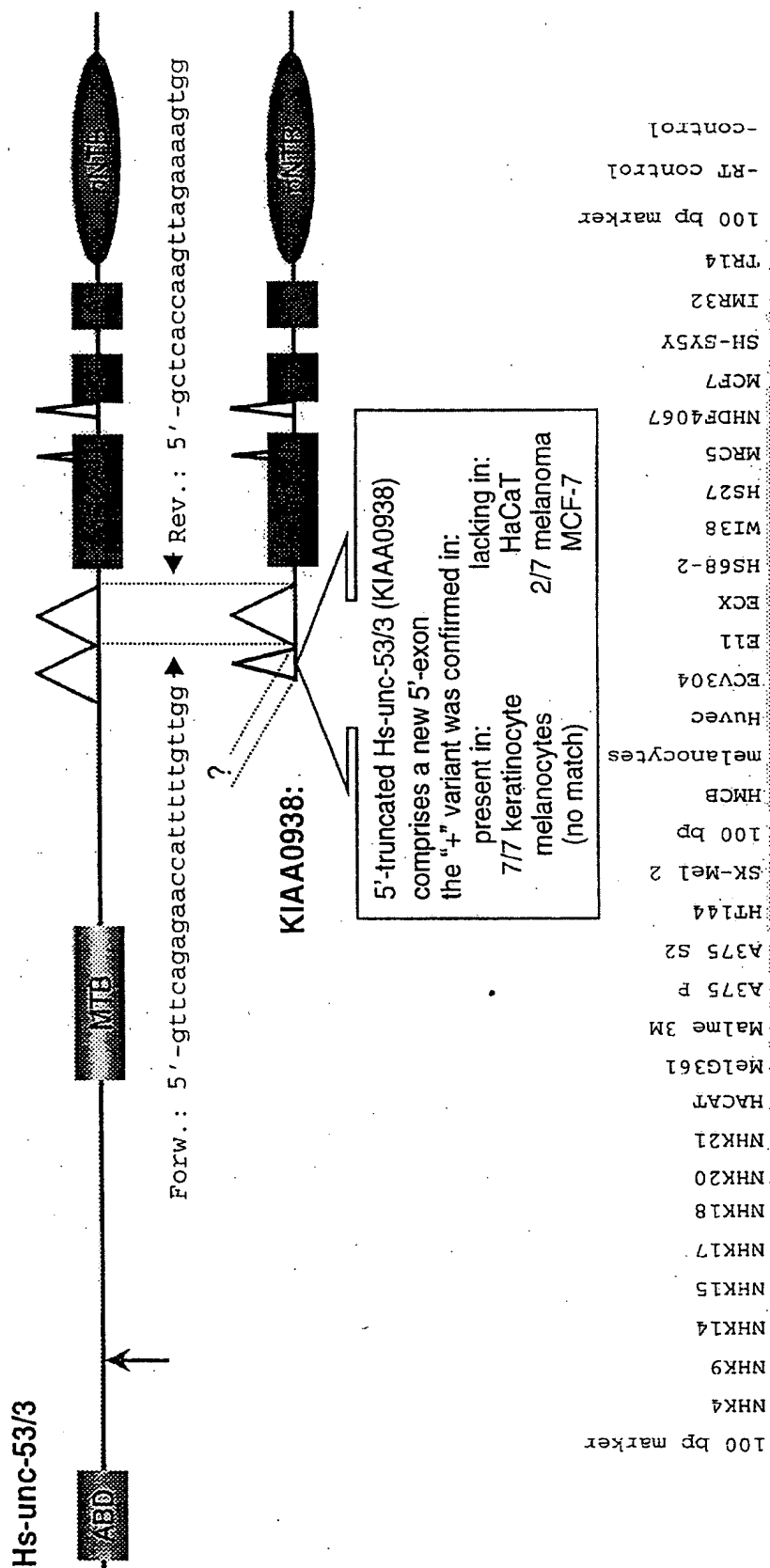
FIG. 6b Hs-unc-53/2 expression: RT-PCR studies*



(*) cDNA quality was assessed using a Hs-ARPP0 PCR reaction: all cDNAs contained the 650 bp Hs-ARPP0 fragment

20/56

FIG. 6c Expression of AB023155 (KIAA0938)



(*) cDNA quality was assessed using a Hs-ARPP0 PCR reaction: all cDNAs contained the Hs-ARPP0 fragment

27/56

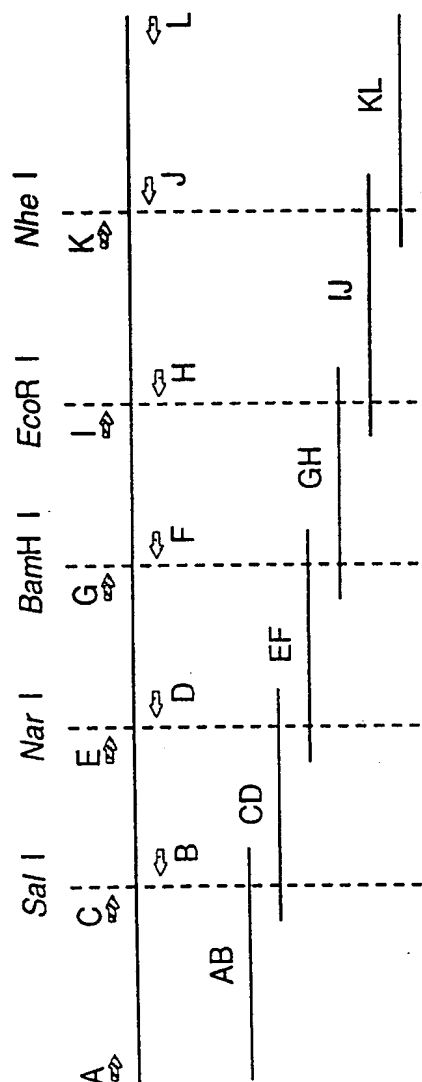
16	1..120	UNES3_KIAA_PCR01	10	20	30	40	50	60	70	80	90	100	110	120
1	1..120	UNES3_KIAA_PCR02	10	20	30	40	50	60	70	80	90	100	110	120
2	1..120	UNES3_KIAA_PCR03	10	20	30	40	50	60	70	80	90	100	110	120
3	1..120	UNES3_KIAA_PCR04	10	20	30	40	50	60	70	80	90	100	110	120
4	1..120	UNES3_KIAA_PCR05	10	20	30	40	50	60	70	80	90	100	110	120
5	1..120	UNES3_KIAA_PCR06	10	20	30	40	50	60	70	80	90	100	110	120
6	1..120	UNES3_KIAA_PCR07	10	20	30	40	50	60	70	80	90	100	110	120
7	1..120	UNES3_KIAA_PCR08	10	20	30	40	50	60	70	80	90	100	110	120
8	1..120	UNES3_KIAA_PCR09	10	20	30	40	50	60	70	80	90	100	110	120
9	1..120	UNES3_KIAA_PCR10	10	20	30	40	50	60	70	80	90	100	110	120
10	1..120	UNES3_KIAA_PCR11	10	20	30	40	50	60	70	80	90	100	110	120
11	1..120	UNES3_KIAA_PCR12	10	20	30	40	50	60	70	80	90	100	110	120
12	1..120	UNES3_KIAA_PCR13	10	20	30	40	50	60	70	80	90	100	110	120
13	1..120	UNES3_KIAA_PCR14	10	20	30	40	50	60	70	80	90	100	110	120
14	1..120	UNES3_KIAA_PCR15	10	20	30	40	50	60	70	80	90	100	110	120
15	1..120	UNES3_KIAA_PCR16	10	20	30	40	50	60	70	80	90	100	110	120
16	1..120	UNES3_KIAA_PCR17	10	20	30	40	50	60	70	80	90	100	110	120
17	1..120	UNES3_KIAA_PCR18	10	20	30	40	50	60	70	80	90	100	110	120
18	1..120	UNES3_KIAA_PCR19	10	20	30	40	50	60	70	80	90	100	110	120
19	1..120	UNES3_KIAA_PCR20	10	20	30	40	50	60	70	80	90	100	110	120
20	1..120	UNES3_KIAA_PCR21	10	20	30	40	50	60	70	80	90	100	110	120
21	1..120	UNES3_KIAA_PCR22	10	20	30	40	50	60	70	80	90	100	110	120
22	1..120	UNES3_KIAA_PCR23	10	20	30	40	50	60	70	80	90	100	110	120
23	1..120	UNES3_KIAA_PCR24	10	20	30	40	50	60	70	80	90	100	110	120
24	1..120	UNES3_KIAA_PCR25	10	20	30	40	50	60	70	80	90	100	110	120
25	1..120	UNES3_KIAA_PCR26	10	20	30	40	50	60	70	80	90	100	110	120
26	1..120	UNES3_KIAA_PCR27	10	20	30	40	50	60	70	80	90	100	110	120
27	1..120	UNES3_KIAA_PCR28	10	20	30	40	50	60	70	80	90	100	110	120
28	1..120	UNES3_KIAA_PCR29	10	20	30	40	50	60	70	80	90	100	110	120
29	1..120	UNES3_KIAA_PCR30	10	20	30	40	50	60	70	80	90	100	110	120
30	1..120	UNES3_KIAA_PCR31	10	20	30	40	50	60	70	80	90	100	110	120
31	1..120	UNES3_KIAA_PCR32	10	20	30	40	50	60	70	80	90	100	110	120
32	1..120	UNES3_KIAA_PCR33	10	20	30	40	50	60	70	80	90	100	110	120
33	1..120	UNES3_KIAA_PCR34	10	20	30	40	50	60	70	80	90	100	110	120
34	1..120	UNES3_KIAA_PCR35	10	20	30	40	50	60	70	80	90	100	110	120
35	1..120	UNES3_KIAA_PCR36	10	20	30	40	50	60	70	80	90	100	110	120
36	1..120	UNES3_KIAA_PCR37	10	20	30	40	50	60	70	80	90	100	110	120
37	1..120	UNES3_KIAA_PCR38	10	20	30	40	50	60	70	80	90	100	110	120
38	1..120	UNES3_KIAA_PCR39	10	20	30	40	50	60	70	80	90	100	110	120
39	1..120	UNES3_KIAA_PCR40	10	20	30	40	50	60	70	80	90	100	110	120
40	1..120	UNES3_KIAA_PCR41	10	20	30	40	50	60	70	80	90	100	110	120
41	1..120	UNES3_KIAA_PCR42	10	20	30	40	50	60	70	80	90	100	110	120
42	1..120	UNES3_KIAA_PCR43	10	20	30	40	50	60	70	80	90	100	110	120
43	1..120	UNES3_KIAA_PCR44	10	20	30	40	50	60	70	80	90	100	110	120
44	1..120	UNES3_KIAA_PCR45	10	20	30	40	50	60	70	80	90	100	110	120
45	1..120	UNES3_KIAA_PCR46	10	20	30	40	50	60	70	80	90	100	110	120
46	1..120	UNES3_KIAA_PCR47	10	20	30	40	50	60	70	80	90	100	110	120
47	1..120	UNES3_KIAA_PCR48	10	20	30	40	50	60	70	80	90	100	110	120
48	1..120	UNES3_KIAA_PCR49	10	20	30	40	50	60	70	80	90	100	110	120
49	1..120	UNES3_KIAA_PCR50	10	20	30	40	50	60	70	80	90	100	110	120
50	1..120	UNES3_KIAA_PCR51	10	20	30	40	50	60	70	80	90	100	110	120
51	1..120	UNES3_KIAA_PCR52	10	20	30	40	50	60	70	80	90	100	110	120
52	1..120	UNES3_KIAA_PCR53	10	20	30	40	50	60	70	80	90	100	110	120
53	1..120	UNES3_KIAA_PCR54	10	20	30	40	50	60	70	80	90	100	110	120
54	1..120	UNES3_KIAA_PCR55	10	20	30	40	50	60	70	80	90	100	110	120
55	1..120	UNES3_KIAA_PCR56	10	20	30	40	50	60	70	80	90	100	110	120
56	1..120	UNES3_KIAA_PCR57	10	20	30	40	50	60	70	80	90	100	110	120
57	1..120	UNES3_KIAA_PCR58	10	20	30	40	50	60	70	80	90	100	110	120
58	1..120	UNES3_KIAA_PCR59	10	20	30	40	50	60	70	80	90	100	110	120
59	1..120	UNES3_KIAA_PCR60	10	20	30	40	50	60	70	80	90	100	110	120
60	1..120	UNES3_KIAA_PCR61	10	20	30	40	50	60	70	80	90	100	110	120
61	1..120	UNES3_KIAA_PCR62	10	20	30	40	50	60	70	80	90	100	110	120
62	1..120	UNES3_KIAA_PCR63	10	20	30	40	50	60	70	80	90	100	110	120
63	1..120	UNES3_KIAA_PCR64	10	20	30	40	50	60	70	80	90	100	110	120
64	1..120	UNES3_KIAA_PCR65	10	20	30	40	50	60	70	80	90	100	110	120
65	1..120	UNES3_KIAA_PCR66	10	20	30	40	50	60	70	80	90	100	110	120
66	1..120	UNES3_KIAA_PCR67	10	20	30	40	50	60	70	80	90	100	110	120
67	1..120	UNES3_KIAA_PCR68	10	20	30	40	50	60	70	80	90	100	110	120
68	1..120	UNES3_KIAA_PCR69	10	20	30	40	50	60	70	80	90	100	110	120
69	1..120	UNES3_KIAA_PCR70	10	20	30	40	50	60	70	80	90	100	110	120
70	1..120	UNES3_KIAA_PCR71	10	20	30	40	50	60	70	80	90	100	110	120
71	1..120	UNES3_KIAA_PCR72	10	20	30	40	50	60	70	80	90	100	110	120
72	1..120	UNES3_KIAA_PCR73	10	20	30	40	50	60	70	80	90	100	110	120
73	1..120	UNES3_KIAA_PCR74	10	20	30	40	50	60	70	80	90	100	110	120
74	1..120	UNES3_KIAA_PCR75	10	20	30	40	50	60	70	80	90	100	110	120
75	1..120	UNES3_KIAA_PCR76	10	20	30	40	50	60	70	80	90	100	110	120
76	1..120	UNES3_KIAA_PCR77	10	20	30	40	50	60	70	80	90	100	110	120
77	1..120	UNES3_KIAA_PCR78	10	20	30	40	50	60	70	80	90	100	110	120
78	1..120	UNES3_KIAA_PCR79	10	20	30	40	50	60	70	80	90	100	110	120
79	1..120	UNES3_KIAA_PCR80	10	20	30	40	50	60	70	80	90	100	110	120
80	1..120	UNES3_KIAA_PCR81	10	20	30	40	50	60	70	80	90	100	110	120
81	1..120	UNES3_KIAA_PCR82	10	20	30	40	50	60	70	80	90	100	110	120
82	1..120	UNES3_KIAA_PCR83	10	20	30	40	50	60	70	80	90	100	110	120
83	1..120	UNES3_KIAA_PCR84	10	20	30	40	50	60	70	80	90	100	110	120
84	1..120	UNES3_KIAA_PCR85	10	20	30	40	50	60	70	80	90	100	110	120
85	1..120	UNES3_KIAA_PCR86	10	20	30	40	50	60	70	80	90	100	110	120
86	1..120	UNES3_KIAA_PCR87	10	20	30	40	50	60	70	80	90	100	110	120
87	1..120	UNES3_KIAA_PCR88	10	20	30	40	50	60	70	80	90	100	110	120
88	1..120	UNES3_KIAA_PCR89	10	20	30	40	50	60	70	80	90	100	110	120
89	1..120	UNES3_KIAA_PCR90	10	20	30	40	50	60	70	80	90	100	110	120
90	1..120	UNES3_KIAA_PCR91	10	20	30	40	50	60	70	80	90	100	110	120
91	1..120	UNES3_KIAA_PCR92	10	20	30	40	50	60	70	80	90	100	110	120
92	1..120	UNES3_KIAA_PCR93	10	20	30	40	50	60	70	80	90	100	110	120
93	1..120	UNES3_KIAA_PCR94	10	20	30	40	50	60	70	80	90	100	110	120
94	1..120	UNES3_KIAA_PCR95	10	20	30	40	50	60	70	80	90	100	110	120
95	1..120	UNES3_KIAA_PCR96	10	20	30	40	50	60	70	80	90	100	110	120
96</															

Figure 6d.

28/56

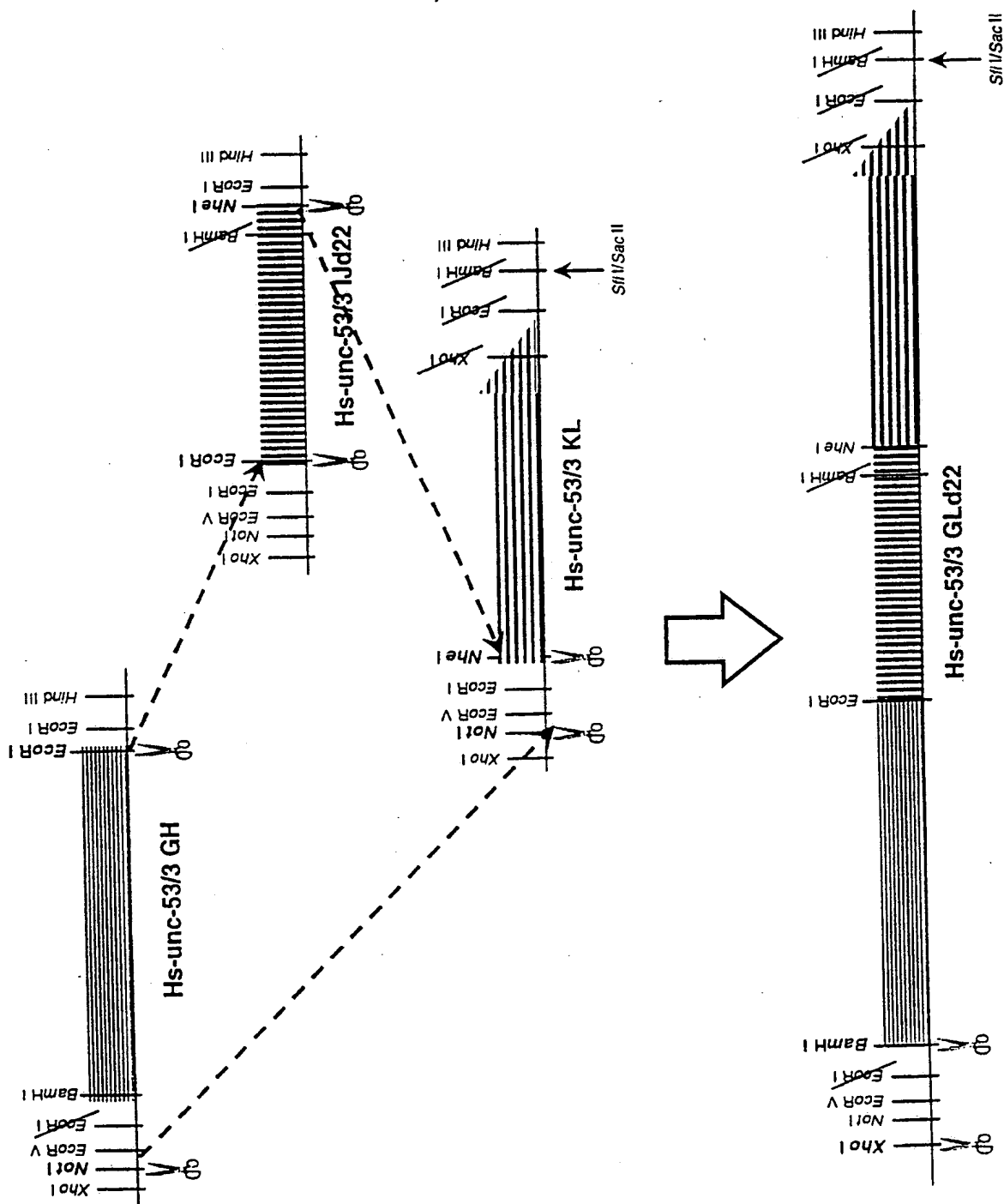
Figure 7a. (1) Strategy for cloning 1-2 kb Hu-unc-53/3 fragments

Schematic:



29/56

Figure 7a. (2) Strategy for cloning the 3' end of Hs-unc-53/3



31/56

Figure 7a. (4) Strategy for cloning the full-length Hs-unc-53/3 construct

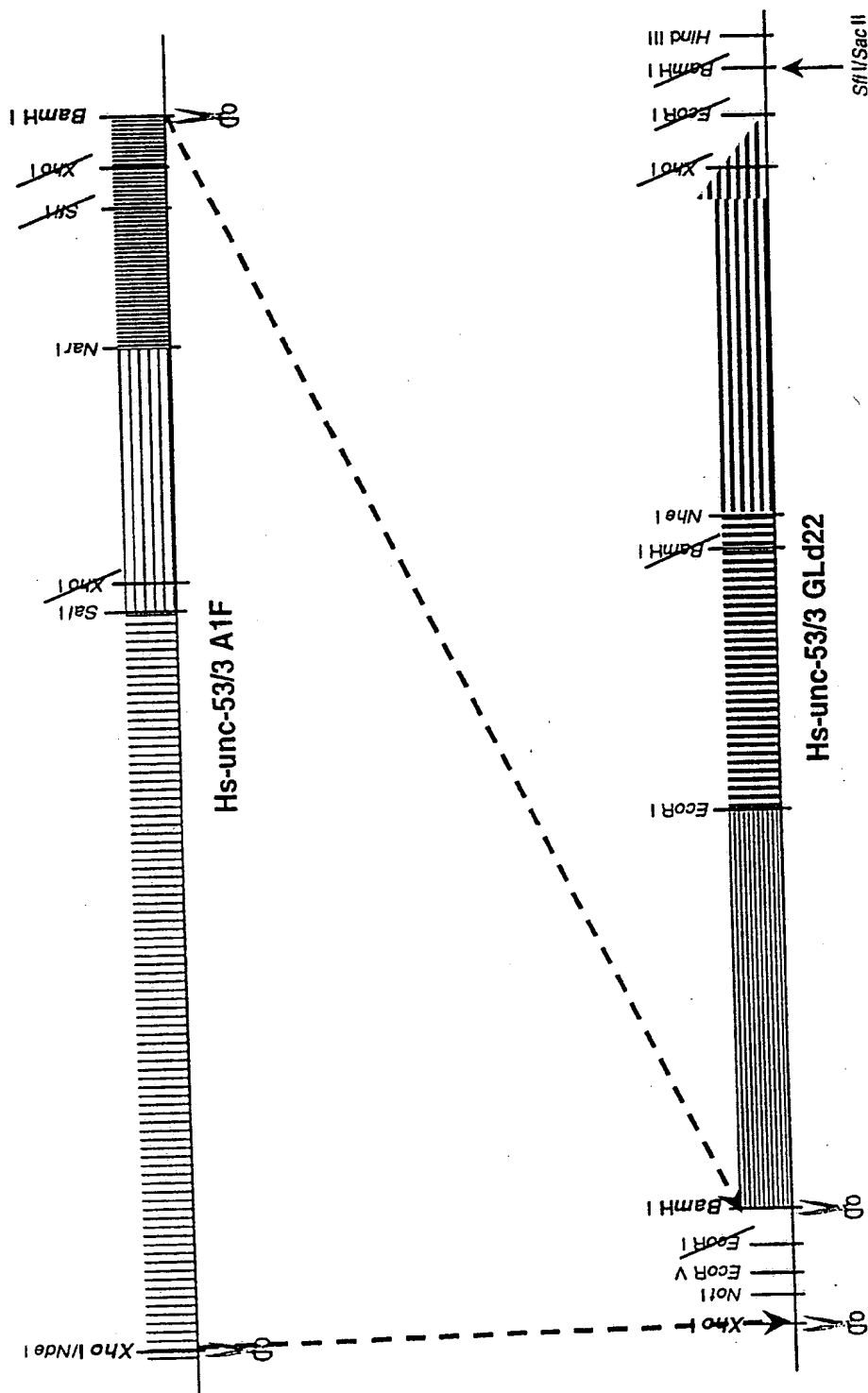
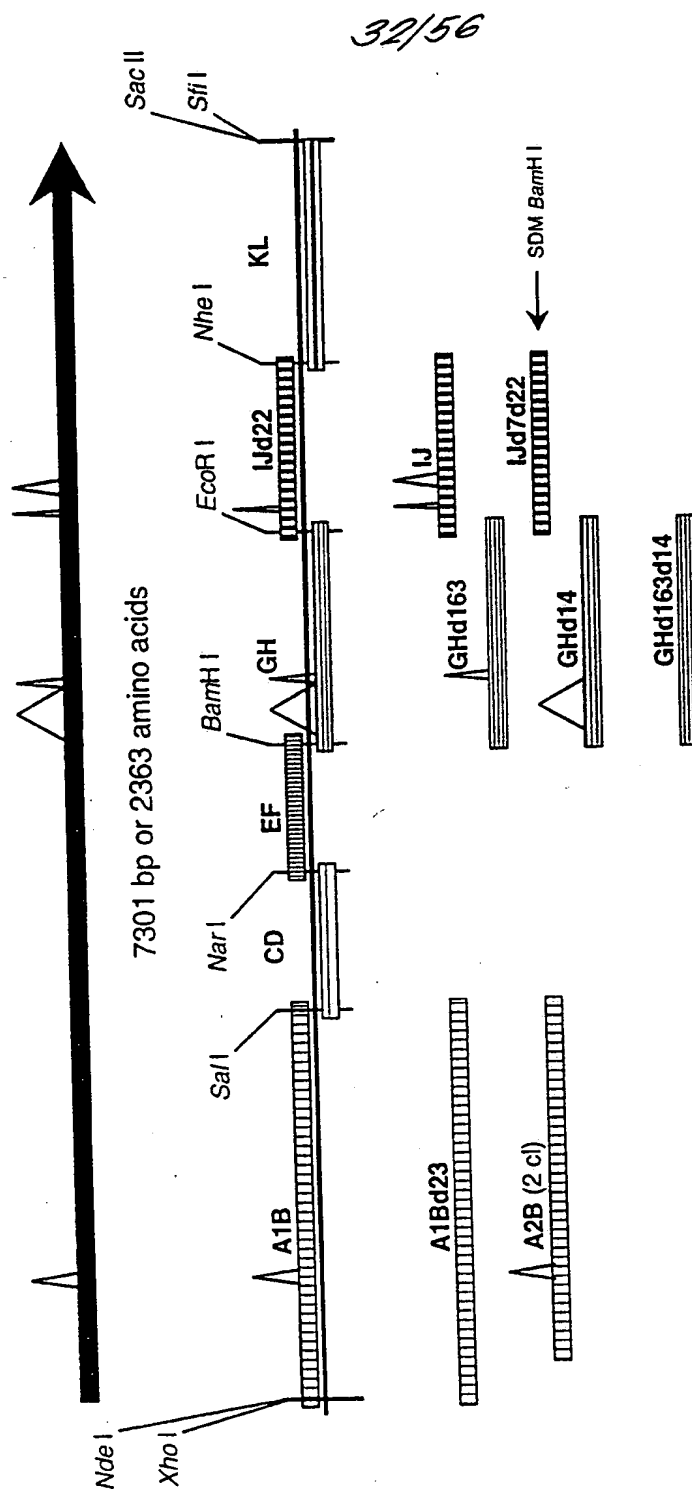
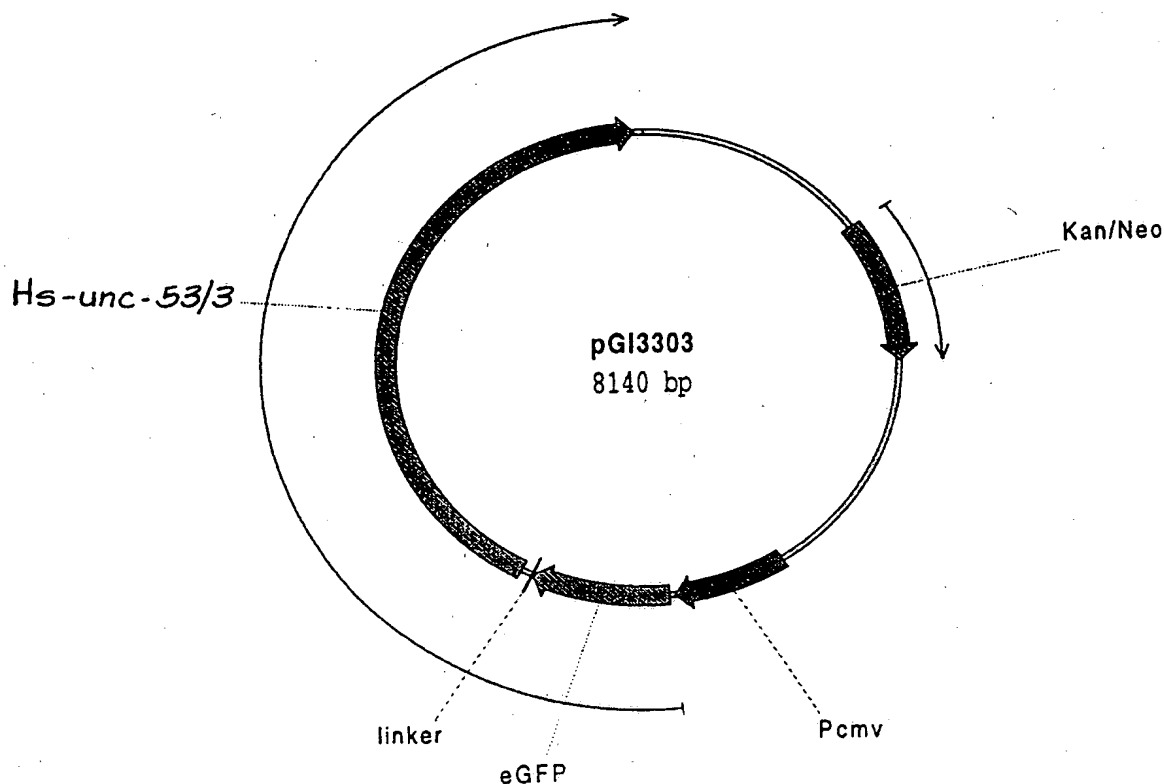


Figure 7a. (5) Cloning of the Hs-unc-53/3-A1L d22 variant



33/56

Figure 7b: Illustration of the plasmid map and the nucleotide sequence of the pGI3303 expression vector (C-terminal Hs-unc-53/3 fragment in fusion with GFP)



ID	pGI3303	circular DNA; 8140 BP				
FT	CDS	1225..2019				
FT		/vntifkey="4"				
FT		/label=Kan/Neo				
FT	CDS	3942..4658				
FT		/vntifkey="4"				
FT		/label=eGFP				
FT	CDS	4719..8102				
FT		/vntifkey="4"				
FT		/label=Hs-unc-53/3				
FT	CDS	4659..4718				
FT		/vntifkey="4"				
FT		/label=linker				
FT	promoter	3330..3918				
FT		/vntifkey="29"				
FT		/label=Pcmv				
SQ	SEQUENCE	8140 BP;				
	CTAGATAACT	GATCATAATC	AGCCATACCA	CATTTGTAGA	GGTTTTACTT	GCTTTAAAAA 60
	ACCTCCCACA	CCTCCCCCTG	AACCTGAAAC	ATAAAATGAA	TGCAATTGTT	GTGTTAACT 120
	TGTTTATTGC	AGCTTATAAT	GGTTACAAAT	AAAGCAATAG	CATCACAAAT	TTCACAAATA 180
	AAGCATTTT	TTCAGTGCAT	TCTAGTTGTG	GTTTGTCCAA	ACTCATCAAT	GTATCTTAAC 240
	GCGTAAATTG	TAAGCGTTAA	TATTTTGTTA	AAATTCGCGT	TAAATTTTTG	TTAAATCAGC 300
	TCATTTTFTA	ACCAATAGGC	CGAAATCGGC	AAAATCCCTT	ATAAATCAAA	AGAATAGACC 360
	GAGATAGGGT	TGAGTGTGT	TCCAGTTTGG	AACAAGAGTC	CACTATTAAA	GAACGTGGAC 420
	TCCAACGTCA	AAGGGCGAAA	AACCGTCTAT	CAGGGCGATG	GCCCACTACG	TGAACCATCA 480
	CCCTAATCAA	GTTTTTTGGG	GTCGAGGTGC	CGTAAAGCAC	TAAATCGGAA	CCCTAAAGGG 540
	AGCCCCCGAT	TTAGAGCTTG	ACGGGGAAAG	CCGGCGAACG	TGGCGAGAAA	GGGAGGGAAG 600
	AAAGCGAAA	GAGCGGGCGC	TAGGGCGCTG	GCAAGTGTAG	CGGTCACGCT	GCGCGTAACC 660
	ACCACACCCG	CCGCGCTTAA	TGCGCCGCTA	CAGGGCGCGT	CAGGTGGCAC	TTTTCGGGGA 720
	AATGTGCGCG	GAACCCCTAT	TTGTTTATTT	TTCTAAATAC	ATTCAAATAT	GTATCCGCTC 780
	ATGAGACAAT	AACCTGATA	AATGCTTCAA	TAATATTGAA	AAAGGAAGAG	TCCTGAGGCG 840

34/56
Figure 7b (CONTINUED 1)

GAAGAACCA	GCTGTGGAAT	GTGTGTCACT	TAGGGTGTGG	AAAGTCCCCA	GGCTCCCCAG	900
CAGGCAGAA	TATGCAAAGC	ATGCATCTCA	ATTAGTCAGC	AACCAGGTGT	GGAAAGTCCC	960
CAGGCTCCCC	AGCAGGCAGA	AGTATGCAAA	GCATGCATCT	CAATTAGTCA	GCAACCATAG	1020
TCCCGCCCCCT	AACTCCGCCC	ATCCCGCCCC	TAACTCCGCC	CAGTTCGGCC	CATTCTCCGC	1080
CCCATGGCTG	ACTAATTTTT	TTTATTTATG	CAGAGGCCGA	GGCCGCCTCG	GCCTCTGAGC	1140
TATTCCAGAA	GTAAGTGAAG	GGCTTTTGTG	GAGGCCTAGG	CTTTTGCAAA	GATCGATCAA	1200
GAGACAGGAT	GAGGATCGTT	TCGCATGATT	GAACAAGATG	GATTGCACGC	AGGTTCTCCG	1260
GCCGCTTGGG	TGGAGAGGCT	ATTCGGCTAT	GACTGGGCAC	AACAGACAAT	CGGCTGTCTT	1320
GATGCCGCCG	TGTTCCGGCT	GTCAAGCCAG	GGGCGCCCGG	TTCTTTTGT	CAAGACCGAC	1380
CTGTCCGGTG	CCCTGAATGA	ACTGCAAGAC	GAGGCAGCGC	GGCTATCGTG	GCTGGCCACG	1440
ACGGGCGTTC	CTTGCGCAGC	TGTGCTCGAC	GTGTGCACTG	AAGCGGGAAG	GGACTGGCTG	1500
CTATTGGGCG	AAGTGCCGGG	GCAGGATCTC	CTGTCACTCT	ACCTTGCTCC	TGCCGAGAAA	1560
GTATCCATCA	TGGCTGATGC	AATGCGGCGG	CTGCATACGC	TTGATCCGGC	TACCTGCCCA	1620
TTCGACCACC	AAGCGAAACA	TCGCATCGAG	CGAGCACGTA	CTCGGATGGA	AGCCGGTCTT	1680
GTCGATCAGG	ATGATCTGGA	CGAAGAGCAT	CAGGGGCTCG	CGCCAGCCGA	ACTGTTCCGC	1740
AGGCTCAAGG	CGAGCATGCC	CGACGGCGAG	GATCTCGTCG	TGACCCATGG	CGATGCCTGC	1800
TTGCCGAATA	TCATGGTGGA	AAATGGCCGC	TTTCTGGAT	TCATCGACTG	TGGCCGGCTG	1860
GGTGTGGCGG	ACCGCTATCA	GGACATAGCG	TTGGCTACCC	GTGATATTGC	TGAAGAGCTT	1920
GGCGGCGAAT	GGGCTGACCG	CTTCTCGTGC	CTTTACGGTA	TGCGCCGCTC	CGATTCCGAG	1980
CGCATCGCCT	TCTATCGCCT	TCTTGACGAG	TTCTTCTGAG	CGGGACTCTG	GGGTTCCGAA	2040
TGACCGACCA	AGCGACGCCC	AACTTGCCAT	CACGAGATTT	CGATTCCACC	GCCGCTTCTT	2100
ATGAAAGGTT	GGGCTTCGGA	ATCGTTTTC	GGGACGCCGG	CTGGATGATC	CTCCAGCCGG	2160
GGGATCTCAT	GCTGGAGTTC	TTCCGCCACC	CTAGGGGGAG	GCTAATGAA	ACACGGAAGG	2220
AGACAATACC	GGAAGGAACC	CGCGCTATGA	CGGCAATAAA	AAGACAGAAT	AAAACGCACG	2280
GTGTTGGGTC	GTTTGTTCAT	AAACGCGGGG	TTCCGTCCCA	GGGCTGGCAC	TCTGTGATA	2340
CCCCACCGAG	ACCCATTGG	GGCCAATACG	CCCGCGTTTC	TTCTTTTTC	CCACCCACC	2400
CCCCAAGTTC	GGGTGAAGGC	CCAGGGCTCG	CAGCCAAAGT	CGGGGCGGCA	GGCCCTGCCA	2460
TAGCCTCAGG	TTACTCATAT	ATACTTTAGA	TTGATTTAAA	ACTTCATTTT	TAATTTAAAA	2520
GGATCTAGGT	GAAGATCCTT	TTTGATAATC	TCATGACCAA	AATCCCTTAA	CGTGAGTTTT	2580
CGTTCCTACTG	AGCGTCAGAC	CCCGTAGAAA	AGATCAAAGG	ATCTTCTTGA	GATCCTTTTT	2640
TTCTGCGCGT	AATCTGCTGC	TTGCAAACAA	AAAAACCACC	GCTACCAGCG	GTGGTTTGT	2700
TGCCGGATCA	AGAGCTACCA	ACTCTTTTTC	CGAAGGTAAC	TGGCTTCAGC	AGAGCGCAGA	2760
TACCAAATAC	TGTCTTCTA	GTGTAGCCGT	AGTTAGGCCA	CCACTTCAAG	AACTCTGTAG	2820
CACCGCCTAC	ATACCTCGCT	CTGCTAATCC	TGTTACCAGT	GGCTGCTGCC	AGTGGCGATA	2880
AGTCGTGTCT	TACCGGGTTG	GACTCAAGAC	GATAGTTACC	GGATAAGGCG	CAGCGGTCCG	2940
GCTGAACGGG	GGGTTCTGTC	ACACAGCCCA	GCTTGGAGCG	AACGACCTAC	ACCGAACTGA	3000
GATACCTACA	CGGTGAGCTA	TGAGAAAGCG	CCACGCTTCC	CGAAGGGAGA	AAGGCGGACA	3060
GGTATCCGGT	AAGCGGCAGG	GTCCGAACAG	GAGAGCGCAC	GAGGGAGCTT	CCAGGGGGAA	3120
ACGCCTGGTA	TCTTTATAGT	CCTGTGCGGT	TTCCGCCACT	CTGACTTGAG	CGTCGATTTT	3180
TGTGATGCTC	GTCAGGGGGG	CGGAGCCTAT	GGAAAAACGC	CAGCAACGCG	GCCTTTTTAC	3240
GGTTCCTGGC	CTTTTGCTGG	CCTTTGCTC	ACATGTTCTT	TCCTGCGTTA	TCCCTGTGAT	3300
CTGTGGATAA	CCGTATTACC	GCCATGCATT	AGTTATTAAT	AGTAATCAAT	TACGGGGTCA	3360
TTAGTTTCATA	GCCCATATAT	GGAGTTCCCG	GTTACATAAC	TTACGGTAAA	TGGCCCGCCT	3420
GGCTGACCGC	CCAACGACCC	CCGCCATTG	ACGTCAATAA	TGACGTATGT	TCCCATAGTA	3480
ACGCCAATAG	GGACTTTCCA	TTGACGTCAA	TGGGTGGAGT	ATTACGGTA	AACTGCCCAC	3540
TTGGCAGTAC	ATCAAGTGTA	TCATATGCCA	AGTACGCCCC	CTATTGACGT	CAATGACGGT	3600
AAATGGCCCC	CCTGGCATT	TGCCCCAGTAC	ATGACCTTAT	GGGACTTCC	TACTTGGCAG	3660
TACATCTACG	TATTAGTCAT	CGCTATTACC	ATGGTGATGC	GGTTTTGGCA	GTACATCAAT	3720
GGCGGTGGAT	AGCGGTTTGA	CTCACGGGGA	TTTCCAAAGT	TCCACCCCAT	TGACGTCAAT	3780
GGGAGTTTGT	TTTGGCACCA	AAATCAACGG	GACTTTCCAA	AATGTCGTAA	CAACTCCGCC	3840
CCATTGACGC	AAATGGGCGG	TAGGCGTGTA	CGGTGGGAGG	TCTATATAAG	CAGAGCTGGT	3900
TTAGTGAACC	GTCAGATCCG	CTAGCGCTAC	CGGTGCGCAC	CATGGTGAGC	AAGGGCGAGG	3960
AGTGTGTTAC	CGGGGTGGTG	CCCATCTGCG	TCGAGCTGGA	CGGCGACGTA	AACGGCCACA	4020
AGTTACAGCT	GTCGGGCGAG	GGCGAGGGCG	ATGCCACCTA	CGGCAAGCTG	ACCCGGAAGT	4080
TCATCTGCAC	CACCGGCAAG	CTGCCCCGTC	CCTGGCCAC	CCTCGTGACC	ACCCTGACCT	4140
ACGGCGTGCA	GTGCTTCAGC	CGCTACCCCG	ACCACATGAA	GCAGCACGAC	TTCTTCAAGT	4200
CCGCCATGCC	CGAAGGCTAC	GTCCAGGAGC	GCACCATCTT	CTTCAAGGAC	GACGGCAACT	4260
ACAAGACCCG	CGCCGAGGTG	AAGTTCGAGG	GCGACACCCT	GGTGAACCGC	ATCGAGCTGA	4320
AGGGCATCGA	CTTCAAGGAG	ACGGGCAACA	TCCTGGGGGA	CAAGCTGGAG	TACAACCTACA	4380
ACAGCCACAA	CGTCTATATC	ATGGCCGACA	AGCAGAAGAA	CGGCATCAAG	GTGAACCTTCA	4440
AGATCCGCCA	CAACATCGAG	GACGGCAGCG	TGCAGCTCGC	CGACCACTAC	CAGCAGAACA	4500
CCCCCATCGG	CGACGGCCCC	GTGCTGTGTC	CCGACAACCA	CTACCTGAGC	ACCCAGTCCG	4560
CCCTGAGCAA	AGACCCCAAC	GAGAAGCGCG	ATCACATGGT	CCTGCTGGAG	TTCGTGACCG	4620
CCGCCGGGAT	CACTCTCGGC	ATGGACGAGC	TGTACAAGTC	CGGACTCAGA	TCTCGAGCTC	4680
AAGTTCGAA	TTCTGCACTC	GACGGTACCG	CGGGCCCGGG	ATCCAAGTAT	CCAGATATTG	4740
CCTCACCCAC	ATTTCGAAGG	TTGTTTGGTG	CCAAGGCAGG	TGGCAAAATCT	GCCTCTGCAC	4800
CTAATACTGA	GGGTGTGAAA	TCTTCTCAG	TAATGCCCAG	CCCTAGTACC	ACATTAGCGC	4860
GGCAAGGCAG	TCTGGAGTCA	CCGTCTGTCG	GTACGGGCAG	CATGGGCAGT	GCTGGTGGGC	4920
TAAGCGGCAG	CAGCAGCCCT	CTCTTCAATA	AACCCTCAGA	CTTAACCTACA	GATGTTATAA	4980
GCTTAAGTCA	CTCGTTGGCC	TCCAGCCAG	CATCGGTTCA	CTCTTTCACA	TCAGGTGGTC	5040
TCGTGTGGGC	TGCCAATATG	AGCAGTTTCT	CTGCAGGCAG	CAAGGATACT	CCGAGCTACC	5100
AGTCCATGAC	TAGCCTCCAC	ACGAGCTCTG	AGTCCATTGA	CCTCCCCCTC	AGCCATCATG	5160
GCTCCTTGTG	TGGACTGACC	ACAGGCTCCA	ACGAGGTCCA	GAGCCTGCTC	ATGAGAACGG	5220
GTAGTGTGAG	ATCTACTCTC	TCAGAAAGCA	TGCAGCTTGA	CAGAAATACA	CTACCCAAAA	5280
AGGGACTAAG	ATATACCCCA	TCATCTCGGC	AGGCCAACCA	AGAAGAGGGC	AAAGAGTGGT	5340
TGCGTTCTCA	TTCTACTGGA	GGGCTTCAGG	ACACTGGCAA	CCAGTCACCT	CTGGTTTCCC	5400

35/56

Figure 7b (CONTINUED 2)

CTTCTGCCAT	GTCATCTTCT	GCAGCTGGAA	AATACCACTT	TTCTAACTTG	GTGAGCCCAA	5460
CAAATTTGTC	TCAATTTAAC	CTTCCCGGGC	CCAGCATGAT	GCGCTCAAAC	AGCATCCCAG	5520
CCCAAGACTC	TTCCTTCGAT	CTCTATGATG	ACTCCCAGCT	TTGTGGGAGT	GCCACTTCTC	5580
TGGAGGAAAG	ACCTCGTGCC	ATCAGTCATT	CGGGCTCATT	CAGAGACAGC	ATGGAAGAAG	5640
TTTATGGCTC	TTCATTATCA	CTGGTGTCCA	GCACCTTCTC	TCTTTACTCT	ACAGCTGAAG	5700
AAAAGGCTCA	TTCAGAGCAA	ATCCATAAAC	TGCGGAGAGA	GCTGGTTGCA	TCACAAGAAA	5760
AAGTTGCTAC	CCTCACATCT	CAGCTTTCAG	CAAATGTCTA	CCTTGTAAGC	GCTTTTGAAA	5820
AGAGCTTAGG	GAATATGACT	GGCCGATTGC	AAAGTCTAAC	TATGACAGCG	GAACAAAAGG	5880
AATCTGAACT	TATAGAAGTA	AGAGAAAACCA	TTGAAATGCT	GAAGGCTCAG	AATTCTGCTG	5940
CCCAGGCGGC	TATTCAGGGA	GCACTGAATG	GTCCAGACCA	TCCTCCCAA	GATCTTCGCA	6000
TCAGAAGACA	GCATTCTCT	GAAAGTGTTT	CTAGTATCAA	CAGTGCCACA	AGCCATTCCA	6060
GTATTGGCAG	TGGTAATGAT	GCCGACTCCA	AGAAGAAGAA	AAAGAAAAC	TGGGTGAAC	6120
CTAGAGGAAG	TGAGCTGAGA	AGTTCTTTCA	AACAAGCCTT	TGGGAAGAAA	AAGTCCACCA	6180
AGCCTCCTTC	ATCACATTCT	GACATTGAAG	AGCTTACTGA	TTTATCCCTT	CCGGCATCCC	6240
CCAAGTTACC	CCATAATGCT	GGTGACTGTG	GCTCAGCATC	CATGAAGCCC	TCACAATCTG	6300
CTTCAGCGAT	CTGTGAATGC	ACAGAAGCTG	AGGCAGAGAT	AATCTGTCAG	CTGAAGAGCG	6360
AGCTCAGAGA	AAAGGAATTA	AAATTAACGG	ATATTCGGCT	GGAGGCCCTC	AGCTCTGCTC	6420
ATCATCTTGA	TCAGATCCGG	GAAGCCATGA	ACCGGATGCA	GAATGAAAT	GAAATACTGA	6480
AAGCTGAAAA	TGACCGGTTG	AAGGCAGAAA	CTGGTAACAC	AGCTAAGCCT	ACTCGGCCAC	6540
CGTCAGAATC	CTCAAGCAGC	ACCTCCTCTT	CATCTCCAG	GCAGTCATTA	GGACTTTCTC	6600
TAAACAATTT	GAACATCACA	GAGGCTGTTA	GCTCAGATAT	TTTGCTAGAT	GATGCTGGTG	6660
ATGCAACTGG	ACATAAAGAT	GGCCGCAAGT	TGAAAATTAT	AGTCTCCATA	AGCAAGGGCT	6720
ATGGTCGAGC	AAAGGACCAA	AAATCTCAGG	CATATTTGAT	AGGCTCCATT	GGTGTAGTG	6780
GAAAAACCAA	GTGGGATGTC	TTAGATGGTG	TAATAAGACG	TCTCTTTAAG	GAATATGTAT	6840
TCCGAATTGA	TACATCCACT	AGCCTTGGTC	TGAGCTCTGA	CTGCATTGCT	AGCTACTGTA	6900
TAGGAGACTT	AATTAGATCC	CATAACCTAG	AAGTGCCCTG	ATTGCTGCCT	TGTGGATACC	6960
TTGTTGGAGA	TAATAACATC	ATCACTGTGA	ACCTCAAAGG	GGTAGAAGAA	AATAGTTTGG	7020
ACAGTTTGTG	TTTTGATACG	CTGATTCCTA	AACCAATTAC	CCAAAGGTAC	TTTAACTTGT	7080
TGATGGAGCA	TCACAGAATT	ATACTCTCAG	GACCGAGTGG	TACTGGAAAG	ACCTATTTGG	7140
CAACACAACT	TGCTGAATAT	GTAATAACCA	AATCTGGAAG	GAAAAAACA	GAGGATGCAA	7200
TTGCCACTTT	TAATGTGGAC	CACAAGTCAA	GTAAGGAATT	GCAACAATAT	CTAGCTAAC	7260
TGGCTGAACA	GTGCACTGCT	GATAATAATG	GAGTGGAGCT	CCCAGTTGTA	ATAATTCTTG	7320
ATAATCTTCA	TCATGTGGGC	TCTCTGAGTG	ATATCTTCAA	TGGTTTTCTC	AATTGTAAAT	7380
ACAACAAATG	TCCATATATT	ATTGGAACAA	TGAATCAGGG	AGTTTCTTCA	TCACCAAATC	7440
TAGAGCTGCA	TCACAATTTT	AGGTGGGTAT	TATGTGCAAA	TCATACAGAA	CCAGTGAAAG	7500
GCTTTTTTAG	CAGATATCTT	CGAAGAAAAC	TCATAGAGAT	AGAAATTGAA	AGGAACATTC	7560
GCAATAATGA	CCTAGTCAAA	ATTATAGATT	GGATTCCGAA	GACGTGGCAT	CATCTCAACA	7620
GTTTTTTTGA	AACACACAGT	TCTTCTGACG	TTACCATTGG	TCCCGGACTA	TTCTTCTCTT	7680
GCCCCATGGA	TGTAGAAGGT	TCTAGAGTAT	GGTTCATGGA	TCTCTGGAAC	TATTCTTTAG	7740
TACCTTATAT	TCTGGAGGCA	GTGAGAGAGG	GTCTTCAGAT	GTATGGGAAA	CGCACACCAT	7800
GGGAAGATCC	TTCAAAGTGG	GTGCTTGACA	CATATCCATG	GAGCTCAGCA	ACTCTGCCTC	7860
AGGAGAGCCC	AGCCTTACTT	CAGCTGCGAC	CAGAAGATGT	TGGGTATGAA	AGCTGCACAT	7920
CCACTAAGGA	AGCCACAACC	TCAAAGCACA	TTCCGCAAAC	TGACACAGAA	GGAGATCCCC	7980
TGATGAATAT	GCTAATGAAA	CTCCAAGAA	CAGCCCAATTA	CTCAAGCACA	CAAAGCTGCG	8040
ACAGCGAAAG	CACCAGCCAC	CATGAAGACA	TTTTGGATTG	ATCTCTTGAA	TCTACCCTCT	8100
AGAGGGTGAA	AGCCGAAATC	CAGCACACTG	GCGGCCGTTA			8140

//

Legend: pGI3303 was obtained by inserting the 3421 bp BamHI/SpeI fragment of the Hs-Unc53/3GLd22_PCR2.1_D02 in a BamHI/XbaI opened pEGFPc1 vector (Clontech Inc.). This plasmid encodes an eGFP protein in fusion with the C-terminal half of Hs-unc-53/3 (last 1128 AA). Arrows indicate the ORFs.

36/56

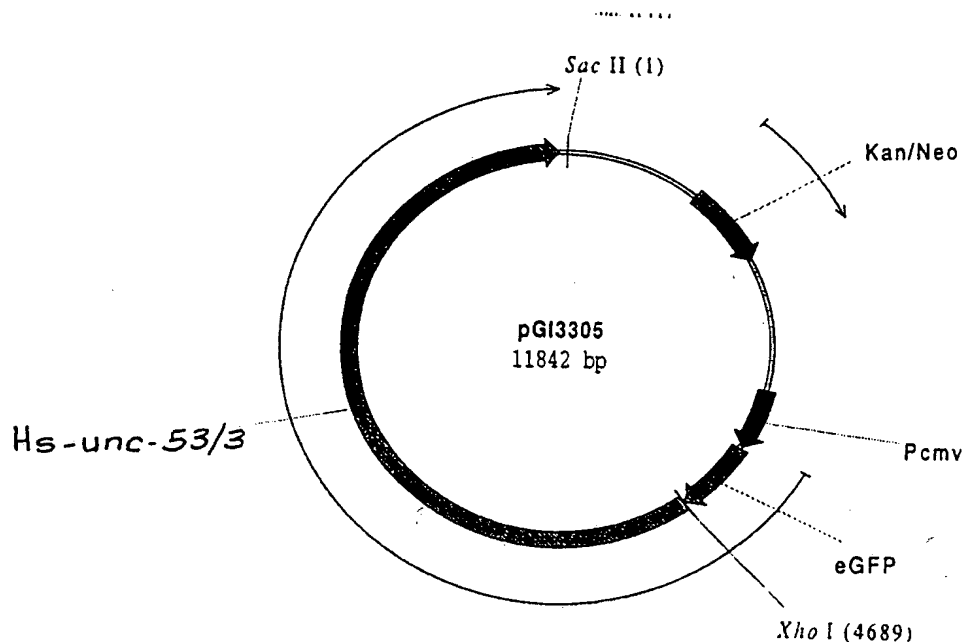
Figure 7c: Illustration of the AA sequence of GFP::C-terminal Hs-unc-53/3 fragment(insert of pGI3303)

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTTKFICTTGKLPVPWPTLVTTLTYGV
QCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNIL
GHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQONTPIGDGPVLLPDNHYLSTQSA
LSKDPNEKRDHMLLEFVTAAGITLGMDELYKSGLRSRAQASNSAVDGTAGPGSKYPDIASPTFRRLFG
AKAGGKSASAPNTEGVKSSSVMPSPSTTLAROGSLESPSSSGTSGMGSAGGLSGSSSPLFNKPSDLTTDV
ISLSHSLASSPASVHSFTSGGLVWAANMSSSSAGSKDTPSYOSMTSLHTSSESIDLPLSHHGSLSGLTT
GTHEVOSLLMRTGSVRSTLSESMOLDRNTLPKKGLRYTPSSROANOEEGKEWLRSHSTGGLODTGNOSP
LVSPSAMSSSAAGKYHFSNLVSPTNLSQFNLPGPSMMRSNSIPAODSSFDLYDDSOLCGSATSLEERPR
AISHSGSFRDSMEEVHGSSLSLVSSTSSLYSTAEKAHSEOIHKLRRELVASOEKVATLTSOLSANAH
VAAFEKSLGNMTGRLOSLTMTAEOKESELIELRETIEMLKAONSAAAQAAIOGALNGPDHPPKDLRIRRO
HSSESVSSINSATSHSSIGSGNDADSKKKKKKNWVNSRGSELRSSFKOAFGKKKSTKPPSSHSDIEELT
DSSLPASPKLPHNAGDCGSASMKPSOSASAICECTEAEAEIIOLKSELREKELKLTDIRLEALSSAHH
LDOIREAMNRMONEIEILKAENDRLKAETGNTAKPTRPPSESSSSTSSSSSRSLGLSLNNLNITEAVS
SDILLDDAGDATGHKDGRSVKIIVSISKGYGRAKDOKSOAYLIGSIGVSGTKWDVLDGVIRRLFKEYV
FRIDTSTSLGLSSDCIASYCIGDLIRSHNLEVPELLPCGYLVGDNNIITVNLKGVEENSLDSFVFDTLI
PKPITORYFNLIMEHHRIILSGPSGTGKTYLANKLAEYVITKSGRKKTEDAIATFNVDHKSSKELOOYL
ANLAEQCSADNNGVELPVVIILDNLHHVGSLSDFNGFLNCKYNKCPYIIGTMNOGVSSSPNLELHNF
RWVLCANHTEPVKGFLGRYLRRKLIEIEIERNIRNNDLVKIIDWIPKTWHHLNSFLETHSSSDVTIGPR
LFLPCPMDVEGSRVWFMDLWNYSLVPYILEAVREGLOMYGKRTPWEDPSKWVLDTPWSSATLPOESPA
LLQLRPEDVGYESCTSTKEATTSKHIPOTDTEGDPLMNMMLKLOEAANYSSTOSCDSESTSHHEDILDS
SLESTL

Legend: Single underlined AA sequence represents eGFP.
 Double underlined AA sequence represents the C-terminal
 fragment of Hs-unc-53/3

37/56

Figure 7d: Illustration of the plasmid map and the nucleotide sequence of the pGI3305 expression vector (full length Hs-unc-53/3 in fusion with GFP)



ID pGI3305 circular DNA; 11842 BP
 FT CDS 1245..2039
 FT /vntifkey="4"
 FT /label=Kan/Neo
 FT CDS 3895..10983
 FT /vntifkey="4"
 FT /label=hHs-unc-53/3\ (full\length)
 FT CDS 3962..4678
 FT /vntifkey="4"
 FT /label=eGFP
 FT promoter 3350..3938
 FT /vntifkey="29"
 FT /label=Pcmv

SQ SEQUENCE 11842 BP;
 GGGCCCGGGA TCCACCGGAT CTAGATAACT GATCATAATC AGCCATACCA CATTGTGTAGA 60
 GGTTTTACTT GCTTTAAAAA ACCTCCCACA CCTCCCCCTG AACCTGAAAC ATAAAAATGAA 120
 TGCAATTGTT GTTGTAACT TGTATTATGC AGCTTATAAT GGTTACAAAT AAAGCAATAG 180
 CATCACAAAT TTCACAAATA AAGCATTTTT TTCACTGCAT TCTAGTTGTG GTTTGTCCAA 240
 ACTCATCAAT GTATCTTAAC GCGTAAATTG TAAGCGTTAA TATTTTGTGA AAATTCGCGT 300
 TAAATTTTGT TTAATCAGC TCATTTTFTA ACCAATAGGC CGAAATCGGC AAAATCCCTT 360
 ATAAATCAAA AGAATAGACC GAGATAGGGT TGAGTGTTGT TCCAGTTTGG AACAAGAGTC 420
 CACTATTAAA GAACGTGGAC TCCAACGTCA AAGGGCGAAA AACCGTCTAT CAGGGCGATG 480
 GCCCACACTACG TGAACCATCA CCCTAATCAA GTTTTTTGGG GTCGAGGTGC CGTAAAGCAC 540
 TAAATCGGAA CCCTAAAGGG AGCCCCGAT TTAGAGCTTG ACGGGGAAAAG CCGGCCGAACG 600
 TGGCGAGAAA GGAAGGGAAG AAAGCGAAAAG GAGCGGGCGC TAGGGCGCTG GCAAGTGTAG 660
 CGGTACAGCT GCGCGTAACC ACCACACCCG CCGCGCTTAA TCGCGCGCTA CAGGGCGCGT 720
 CAGGTGGCAC TTTTCGGGGA AATGTGCGCG GAACCCCTAT TTGTTTATTT TTCTAAATAC 780
 ATTCAATAT GTATCCGCTC ATGAGACAAT AACCTGATA AATGCTTCAA TAATATTGAA 840
 AAAGGAAGAG TCCTGAGGCG GAAAGAACCA GCTGTGGAAT GTGTGTCAAT TAGGGTGTGG 900
 AAAGTCCCAA GGCTCCCCAG CAGGCAGAAG TATGCAAAAG ATGCATCTCA ATTAGTCAGC 960
 AACCAGGTGT GGAAAGTCCC CAGGCTCCCC AGCAGGCAGA AGTATGCAAA GCATGCATCT 1020
 CAATTAGTCA GCAACCATAG TCCCCCCCCT AACTCCGCCC ATCCCGCCCC TAACTCCGCC 1080
 CAGTTCGGCC CATTCTCCGC CCCATGGCTG ACTAATTTTT TTTATTTATG CAGAGGCCGA 1140

38/56

Figure 7d (CONTINUED 1)

GGCCGCTCG	GCCTCTGAGC	TATTCCAGAA	GTAGTGAGGA	GGCTTTTTTG	GAGGCCTAGG	1200
CTTTTGCAAA	GATCGATCAA	GAGACAGGAT	GAGGATCGTT	TCGCATGATT	GAACAAGATG	1260
GATTGCACGC	AGGTTCTCCG	GCCGCTTGGG	TGGAGAGGCT	ATTGCGGTAT	GACTGGGCAC	1320
AACAGACAAT	CGGCTGCTCT	GATGCCGCGG	TGTTCCGGCT	GTGAGCGCAG	GGGCGCCCGG	1380
TTCTTTTTGT	CAAGACCGAC	CTGTCCGGTG	CCCTGAATGA	ACTGCAAGAC	GAGGCAGCGC	1440
GGCTATCGTG	GCTGGCCACG	ACGGGCGTTC	CTTGCGCAGC	TGTGCTCGAC	GTTGTCACGT	1500
AAGCGGGAAG	GGACTGGCTG	CTATTGGGCG	AAGTGCCGGG	GCAGGATCTC	CTGTCACTCTC	1560
ACCTTGCTCC	TGCCGAGAAA	GTATCCATCA	TGGCTGATGC	AATGCGGCGG	CTGCATACGC	1620
TTGATCCGGC	TACCTGCCCA	TTCGACCACC	AAGCGAAACA	TCGCATCGAG	CGAGCACGTA	1680
CTCGGATGGA	AGCCGGTCTT	GTCGATCAGG	ATGATCTGGA	CGAAGAGCAT	CAGGGGCTCG	1740
CGCCAGCCGA	ACTGTTTCGCC	AGGCTCAAGG	CGAGCATGCC	CGACGGCGAG	GATCTCGTCG	1800
TGACCCATGG	CGATGCTGCT	TTGCCGAATA	TCATGGTGGA	AAATGGCCGC	TTTTCTGGAT	1860
TCATCGACTG	TGGCCGGCTG	GGTGTGGCGG	ACCGCTATCA	GGACATAGCG	TGGGCTACCC	1920
GTGATATTGC	TGAAGAGCTT	GGCGGCGAAT	GGGCTGACCG	CTTCCTCGTG	CTTTACGGTA	1980
TGCGCGCTCC	CGATTTCGAG	CGCATCGCCT	TCTATCGCCT	TCTTGACGAG	TCTTCTGAG	2040
CGGGACTCTG	GGGTTTCGAA	TGACCGACCA	AGCGACGCCC	AACCTGCCAT	CACGAGATTT	2100
CGATTCCACC	GCCGCTTCT	ATGAAAGGTT	GGGCTTCGGA	ATCGTTTTCC	GGGACGCGCG	2160
CTGGATGATC	CTCCAGCGCG	GGGATCTCAT	GCTGGAGTTC	TTGCCCCACC	CTAGGGGAGG	2220
GGTAACTGAA	ACACGGGAAG	AGACAATACC	GGAAGGAACC	CGCGCTATGA	CGGCAATAAA	2280
AAGACAGAAT	AAAACGCACG	GTGTTGGGTC	GTTTGTTCAT	AAACGCGGGG	TTGCGTCCCA	2340
GGGCTGGCAC	TCTGTGATA	CCCCACCGAG	ACCCCATTTG	GGCCAATACG	CCGCGTTTC	2400
TTCTTTTTC	CCACCCACCC	CCCCAAGTTC	GGGTGAAGGC	CCAGGGCTCG	CAGCCAACGT	2460
CGGGGCGGCA	GGCCCTGCCA	TAGCCTCAGG	TTACTCATAT	ATACTTTAGA	TTGATTTAAA	2520
ACTTCATTTT	TAATTTAAAA	GGATCTAGGT	GAAGATCCTT	TTTGATAATC	TCATGACCAA	2580
AATCCCTTAA	CGTGAGTTTT	CGTTCCACTG	AGCGTCAGAC	CCCGTAGAAA	AGATCAAAGG	2640
ATCTTCTTGA	GATCCTTTTT	TTCTGCGCGT	AATCTGCTGC	TTGCAACAAA	AAAAACCAAC	2700
GCTACGAGCG	GTGGTTTGT	TGCCGGATCA	AGAGCTACCA	ACTCTTTTTC	CGAAGGTAAC	2760
TGGCTTCAGC	AGAGCGCAGA	TACCAAATAC	TGTCCTTCTA	GTGTAGCCGT	AGTTAGGCCA	2820
CCACTTCAAG	AACCTCTGAG	CACCGCTTAC	ATACCTCGCT	CTGCTAATCC	TGTTACCAGT	2880
GGCTGTGTC	AGTGGCGATA	AGTCGTGTCT	TACCGGGTTG	GACTCAAGAC	GATAGTTACC	2940
GGATAAGGCG	CAGCGGTCGG	GCTGAACGGG	GGGTTTCGTG	ACACAGCCCA	GCTTGGAGCG	3000
AACGACCTAC	ACCGAAGTGA	GATACCTACA	GCGTGAGCTA	TGAGAAAGCG	CCACGCTTCC	3060
CGAAGGGAGA	AAGGCGGACA	GGTATCCGGT	AAGCGGCAGG	GTCCGAACAG	GAGAGCGCAC	3120
GAGGGAGCTT	CCAGGGGGAA	ACGCTGGTGA	TCTTTATAGT	CCTGTCCGGT	TTCCGCCACT	3180
CTGACTTGAG	CGTCGATTTT	TGTGATGCTC	GTCAGGGGGG	CGGAGCCTAT	GGAAAAACGC	3240
CAGCAACGCG	GGCTTTTATC	GGTTCCTGGC	CTTTTGCTGG	CCTTTTGCTC	ACATGTTCTT	3300
TCCTGCGTTA	TCCCTGTGAT	CTGTGGATAA	CCGTATTACC	GCCATGCATT	AGTTATTAAT	3360
AGTAATCAAT	TACGGGGTCA	TTAGTTCATA	GCCCATATAT	GGAGTTCGCG	GTTACATAAC	3420
TTACGGTAAA	TGGCCCGCCT	GGCTGACCGC	CCACGACCCC	CCGCCCATTG	ACGTCAATAA	3480
TGACGTATGT	TCCCATAGTA	ACGCCAATAG	GGACTTTCCA	TTGACGTCAA	TGGGTGGAGT	3540
ATTTACGGTA	AAC TGCCAC	TTGGCAGTAG	ATCAAGTGTA	TCATATGCCA	AGTACGCCCC	3600
CTATTGACGT	CAATGACGGT	AAATGGCCCC	CCTGGCATTG	TGCCCCAGTAC	ATGACCTTAT	3660
GGGACTTTCC	TACTTGGCAG	TACATCTACG	TATTAGTTCAT	CGCTATTACC	ATGGTGATGC	3720
GGTTTTTGGA	GTACATCAAT	GGGCGTGGAT	AGCGGTTTGA	CTCAGGGGGA	TTTCCAAGTC	3780
TCCACCCCAT	TGACGTCAAT	GGGAGTTTGT	TTTGGCACCA	AAATCAACGG	GACTTTCCAA	3840
AATGTCGTAA	CAACTCCGCC	CCATTGACGC	AAATGGGCGG	TAGGCGTGTA	CGGTGGGAGG	3900
TCTATATAAG	CAGAGCTGGT	TTAGTGAACC	GTCAGATCCG	CTAGCGCTAC	CGGTGCGCAC	3960
CATGGTGAGC	AAGGGCGAGG	AGCTGTTTAC	CGGGGTGGTG	CCCATCCTGG	TCGAGCTGGA	4020
CGGCGACGTA	AACGGCCACA	AGTTTACGCT	GTCCGGCGAG	GGCGAGGGCG	ATGCCACCTA	4080
CGGCAAGCTG	ACCCTGAAGT	TCATCTGCAC	CACCGGCAAG	CTGCCCCGTG	CCTGGCCCCAC	4140
CCTCGTGACC	ACCCTGACCT	ACGGCGTGCA	GTGCTTCAGC	CGCTACCCCG	ACCACATGAA	4200
GCACACGAC	TTCTTCAAGT	CCGCCATGCC	CGAAGGCTAC	GTCCAGGAGC	GCACCATCTT	4260
CTTCAAGGAC	GACGGCAACT	ACAAGACCCG	CGCCGAGGTG	AAGTTCGAGG	GCGACACCTT	4320
GGTGAACCGC	ATCGAGCTGA	AGGGCATCGA	CTTCAAGGAG	GACGGCAACA	TCCTGGGGCA	4380
CAAGCTGGAG	TACAACTACA	ACAGCCACAA	CGTCTATATC	ATGGCCGACA	AGCAGAAGAA	4440
CGGCATCAAG	GTGAACCTCA	AGATCCGCCA	CAACATCGAG	GACGGCAGCG	TGCAGCTCGC	4500
CGACCACTAC	CAGCAGAACA	CCCCATCGG	CGACGGCCCC	GTGCTGCTGC	CCGACAACCA	4560
CTACCTGAGC	ACCCAGTCCG	CCCTGAGCAA	AGACCCCAAC	GAGAAAGCGG	ATCACATGTT	4620
CCTGCTGGAG	TTGCTGACCG	CCGCCGGGAT	CACCTCTCGG	ATGGACGAGC	TGTACAAGTA	4680
CTCAGATCTC	GAGCATATGC	CTGTTCTTGG	GGTTGCTTCA	AAACTGAGGC	AGCCAGCTGT	4740
TGGGTCAAAG	CCTGTGCATA	CTGCTCTTCC	GATACCAAAT	CTTGGCACTA	CTGGGTTCACA	4800
GCACTGTTCT	TCAAGACCTT	TGGAACCTGC	TGAAACAGAG	AGCTCCATGC	TTTCTTGTCA	4860
GCTTGCCTTA	AAATCAACCT	GTGAATTTGG	AGAGAAGAAA	CCCCCTCCAG	GAAAAAGCCAA	4920
GGAGAAAGAA	GACAGCAAGA	TTTACACTGA	CTGGGCCAAC	CACCTACCTAG	CAAAATCAGG	4980
CCACAAGCCG	CTGATCAAGG	ACTTGCAACA	AGACATTGCA	GATGGAGTAC	TCCTAGCAGA	5040
AATCATCCAG	ATTATTGCAA	ATGAAAAAGT	TGAAGATATC	AATGGATGTC	CTAGAAGTCA	5100
GTCTCAGATG	ATTGAAAATG	TTGATGTCTG	CCTTAGTTTT	CTAGCAGCCA	GAGGGGTAAA	5160
TGTTCAAGGT	CTATCTGCTG	AAGAAAATAAG	AAATGGAAAC	TTAAAAGCCA	TTCTAGGGCT	5220
GTTTTTCAAGT	TTATCTCGCT	ACAAGCAGCA	ACAACACCAT	CAACAACAGT	ACTATCAGTC	5280
CTTGGTGGAA	CTTCAGCAGC	GAGTTACTCA	CGCTTCCCTT	CCATCGGAAG	CCAGCCAGGC	5340
CAAAACCCAG	CAAGATATGC	AGTCCAGTCT	GGCAGCCAGA	TATGCAACTC	AGTCTAATCA	5400
CAGTGGAAAT	GCAACCAAGT	AAAAAAAGCC	TACTAGGCTT	CCAGGGCCCT	CTAGGGTGCC	5460
TGCTGCAGGA	AGCAGCAGCA	AGGTCCAGGG	AGCCTCTAAT	TTAAATAGGA	GAAGTCAGAG	5520
CTTTAAACAG	ATTGACAAAA	ACAAGCCTCC	AAATTATGCA	AATGGAAACG	AAAAAGATTG	5580
CTCCAAAGGA	CCTCAATCGT	CTTCAGGTGT	AAATGGTAAC	GTGCAGCCTC	CCAGTACTGC	5640
TGGGCAGCCT	CCTGCCCTCTG	CCATCCCTTC	TCCAAGTGCC	AGCAAGCCCT	GGCGCAGCAA	5700

39/56

Figure 7d (CONTINUED 2)

GTCCATGAAT	GTCAAACACA	GTGCCACCTC	CACCATGTTG	ACTGTAAAGC	AGTCAAGTAC	5760
AGCCACCTCC	CCCAACCAT	CTTCAGACAG	ACTGAAGCCA	CCTGTCTCAG	AAGGGGTCAA	5820
AACGTCTCCC	TCAGGACAGA	AATCCATGCT	TGAGAAATTC	AAGCTAGTCA	ATGCCCCGAC	5880
TGCTTTACGC	CCCCCGCAGC	CTCCAGTTTC	AGGACCTAGT	GATGGTGGGA	AGGATGATGA	5940
TGCCTTTTCT	GAATCTGGTG	AAATGGAAGG	TTTAAACAGT	GGTCTGAATA	GTGGTGGCTC	6000
AACAAATAGC	AGTCCCAAAG	TGTCACCTAA	GTTGGCCCTT	CCAAAAGCTG	GAAGCAAAAA	6060
TCTCAGCAAT	AAAAAGTCTT	TGCTACAGCC	AAAGGAAAAA	GAAGAAAAAG	ACAGGGACAA	6120
AAATAAAGTT	TGCACTGAAA	AACCAGTCAA	AGAAGAGAAG	GATCAGGTGA	CAGAGATGGC	6180
TCCAAAAAAG	ACCTCCAAAA	TTGCAAGCTT	GATCCCTAAG	GGCAGCAAGA	CAACAGCAGC	6240
TAAGAAGGAA	AGCTTAATTC	CGTCTTCCAG	TGGTATTCCA	AAACCAGGCT	CTAAAGTTCC	6300
AACAGTAAAG	CAAACCATTT	CACCTGGCAG	CACAGCAAGC	AAAGAGTCTG	AGAAATTTCAG	6360
GACTACCAAG	GGGAGCCCTT	CCCAGTCCTT	ATCTAAGCCT	ATAACCATGG	AGAAAGCAAG	6420
TGCTTCTAGT	TGCTCTGCCC	CTTTGGAAGG	AAGGGAAGCT	GGCCAAGCTT	CTCCTTCTGG	6480
TTCTGTATAC	ATGACAGTGG	CACAAAGCAG	TGGGCAGAGC	ACAGGAAATG	GTGCTGTCCA	6540
ACTCCCTCAA	CAGCAGCAAC	ATAGCCACCC	GAATACCGCG	ACAGTGGCAC	CATTCAATTTA	6600
CAGGGCACAT	TCAGAAAAATG	AAGGTACCGC	TTTACCATCG	GCTGACTCCT	GTACCAGTCC	6660
TACAAAGATG	GACTTATCAT	ATAGTAAGAC	TGCTAAGCAG	TGCCTGGAGG	AGATATCTGG	6720
TGAAGACCTT	GAACAAGAA	GAATGAGAAC	AGTTAAAAAC	ATAGCAGACT	TGAGGCAGAA	6780
TTTAGAAGAG	ACTATGTCCA	GTCTTCGTGG	GACTCAGATA	AGCCACAGCA	CCCTGGAGAC	6840
AACATTTGAC	AGCACTGTGA	CAACAGAAAGT	TAATGGAAGG	ACCATACCCA	ACTTGACAAG	6900
TCGACCCACC	CCCATGACCT	GGAGGTTGGG	CCAGGCATGT	CCGCGACTTC	AGGCGGGAGA	6960
TGCTCCCTCC	CTGGGTGCTG	GCTATCCTCG	CAGTGGTACC	AGTCGATTCA	TCCACACAGA	7020
CCCCCAAGG	TTTATGTATA	CCACGCCCTC	CCGTGAGGCT	GCTGTCTCTA	GGCTGGGAAA	7080
CATGTCACAG	ATTGACATGA	GTGAGAAAGC	AAGCAGTGAC	CTGGACATGT	CTTCTGAGGT	7140
CGATGTGGGT	GGATATATGA	GTGATGGTGA	TATCCTTGGG	AAAAAGTCTCA	GGACTGATGA	7200
CATCAACAGT	GGGTACATGA	CAGATGGAGG	ACTTAACCTA	TATACTAGAA	GTCTGAACCG	7260
AATACCAGAC	ACAGCAACTT	CCCGGGACAT	CATCCAGAGA	GGGGTTCACG	ATGTGACAGT	7320
GGATGCAGAC	AGCTGGGATG	ACAGCAGTTC	AGTGAGCAGT	GGTCTCAGTG	ACACCCCTTA	7380
TAACATCAGC	ACTGATGACC	TGAACACCAC	ATCCTCTGTC	AGCTCTTACT	CCAACATCAC	7440
CGTCCCTCT	AGGAAGAATA	CTCAGCTGAG	GACAGATTCA	GAGAAACGCT	CCACCACAGA	7500
CGAGACCTGG	GATAGTCTCT	AGGAACTGAA	AAAACCAGAA	NAAGATTTTG	ACAGCCCATGG	7560
GGATGCTGGT	GGCAAGTGGA	AGACTGTGTC	CTCTGGACTT	CCTGAAGACC	CCGAGAAGGC	7620
AGGGCAGAAA	GCTTCCCTGT	CTGTTTCACA	GACAGGTTC	TGGAGAGAG	GCATGTCTGC	7680
CCAAGGAGGG	GCGCCATCTA	GGCAGAAAAG	TGGAACAAGT	GCACTCAAAA	CACCCGGGAA	7740
AACCGATGAT	GCCAAAGCTT	CTGAGAAAGG	AAAAGCTCCC	CTAAAAGGAT	CATCTCTACA	7800
AAGATCTCCT	TCAGATGCAG	GAAAAAGCAG	TGGAGATGAA	GGGAAAAAGC	CCCCCTCAGG	7860
CATTGGAAGA	TCGACTGCCA	CCAGCTCCTT	TGGCTTTAAG	AAACCAAGTG	GAGTAGGGTC	7920
ATCTGCCATG	ATCACCAGCA	GTGGAGCAAC	CATAACAAAGT	GGCTCTGCAA	CACTGGGTAA	7980
AATTCCAAAA	TCTGCTGCCA	TTGGCGGGAA	GTCAAATGCA	GGGAGAAAAA	CCAGTTTGGA	8040
CGGTTACAG	AATCAGGATG	ATGTTGTGCT	GCATGTTAGC	TCAAAGACTA	CCCTACAATA	8100
TCGCAGCTTG	CCCCGCCCTT	CAAAATCCAG	CACCAAGTGC	ATTCTCTGGAC	GAGGAGGCCA	8160
CAGATCCAGT	ACCAGCAGTA	TTGATTCCAA	CGTCAGCAGC	AAGTCTGCTG	GGGCCACCAC	8220
CTCGAAACTG	AGAGAACCAG	CTAAAATTGG	GTGAGGGCGC	TCAAGTCTCT	TCACCGTCAA	8280
CCAAACAGAC	AAGGAAAAAG	AAAAAGTAGC	AGTCTCAGAT	TCAGAAAGTG	TTTCTTTGTC	8340
AGGTTCCCCC	AAATCCAGCC	CCACCTCTGC	CAGCGCCTGT	GGTGCACAAG	GTCTCAGGCA	8400
GCCAGGATCC	AAGTATCCAG	ATATTGCCCT	ACCCACATTT	CGAAGGTGTT	TTGGTGCCAA	8460
GGCAGGTGGC	AAATCTGCCT	CTGCACCTAA	TACTGAGGGT	GTGAAATCTT	CCTCAGTAAT	8520
GCCCAGCCCT	AGTACCACAT	TAGCGCGGCA	AGGCAGTCTG	GAGTCACCGT	CGTCCGGTAC	8580
GGGCAGCATG	GGCAGTGCTG	GTGGGCTAAG	CGGCAGCAGC	AGCCCTCTCT	TCAATAAACC	8640
CTCAGACTTA	ACTACAGATG	TTATAAGCTT	AAGTCACTCG	TTGGCCTCCA	GCCCAGCATC	8700
GGTTCACTCT	TTACATCATG	GTGGTCTCGT	GTGGGCTGCC	AATATGAGCA	GTTCCTCTGC	8760
AGGCAGCAAG	GATACTCCGA	GCTACCAGTC	CATGACTAGC	CTCCACACGA	GCTCTGAGTC	8820
CATTGACCTC	CCCCCTCAGC	ATCATGGCTC	CTTGTCTGGA	CTGACCACAG	GCACTCACGA	8880
GGTCCAGAGC	CTGCTCATGA	GAACGGGTAG	TGTGAGATCT	ACTCTCTCAG	AAAGCATGCA	8940
GCTTGACAGA	AATACACTAC	CCAAAAAGGG	ACTAAGATAT	ACCCCATCAT	CTCGGCAGGC	9000
CAACCAAGAA	GAGGGCAAAG	AGTGGTTGCG	TTCTCATTTCT	ACTGGAGGGC	TTCAGGACAC	9060
TGGCAACCAG	TCACCTCTGG	TTTCCCCTTC	TGCCATGTCA	TCTTCTCAG	CTGGAAAAATA	9120
CCACTTTTCT	AACTTGGTGA	GCCCAACAAA	TTTGTCTCAA	TTTAACCTTC	CCGGGCCCCG	9180
CATGATGCGC	TCAAACAGCA	TCCCAGCCCC	AGACTCTTCC	TTGCATCTCT	ATGATGACTC	9240
CCAGCTTTGT	GGGAGTGCCA	CTTCTCTGGA	GGAAGACCT	CGTGCCATCA	GTCAATTCGGG	9300
CTCATTCAGA	GACAGCATGG	AAGAAGTTCA	TGGCTCTTCA	TTATCACTGG	TGTCCAGCAC	9360
TTCTTCTCTT	TACTCTACAG	CTGAAGAAAA	GGCTCATTTA	GAGCAAAATC	ATAAACTGGC	9420
GAGAGAGCTG	GTTGCATCAC	AAGAAAAAGT	TGCTACCCTC	ACATCTCAGC	TTTCAGCAAA	9480
TGCTCACCTT	GTAGCAGCTT	TTGAAAAAG	CTTAGGGAAT	ATGACTGGCC	GATTGCAAG	9540
TCTAAGCTATG	ACAGCGGAAC	AAAAGGAATC	TGAACCTATA	GAACCTAAGAG	AAACCATTTA	9600
AATGCTGAAG	GCTCAGAATT	CTGCTGCCCA	GGCGGCTATT	CAGGGAGCAC	TGAATGGTCC	9660
AGACCATCTT	CCCCAAGATC	TTTCGCATCAG	AAGACAGCAT	TCCTCTGAAA	GTGTTTCTAG	9720
TATCAACAGT	GCCACAAGCC	ATTCCAGTAT	TGGCAGTGGT	AATGATGCCG	ACTCCAAGAA	9780
GAAGAAAAAG	AAAAACTGGG	TGAACCTTAG	AGGAAGTGAG	CTGAGAAGTT	CTTTCAAAACA	9840
AGCCTTTGGG	AAGAAAAAGT	CCACCAAGCC	TCCTTCATCA	CATTCTGACA	TTGAAGAGCT	9900
TACTGATTCA	TCCCTTCCGG	CATCCCCCAA	GTTACCCCAT	AATGCTGGTG	ACTGTGGCTC	9960
AGCATCCATG	AAGCCCTCAC	AATCTGCTTC	AGCGATCTGT	GAATGCACAG	AAGCTGAGGC	10020
AGAGATAATT	CTGCAGCTGA	AGAGCGAGCT	CAGAGAAAAAG	GAATTTAAAT	TAACGGATAT	10080
TGGGCTGGAG	GCCCTCAGCT	CTGCTCATCA	TCTTGATCAG	ATCCGGGAAG	CCATGAACCG	10140
GATGCAGAAT	GAATTTGAAA	TACTGAAAGC	TGAAAAATGAC	CGGTTGAAGG	CAGAACTGG	10200
TAACACAGCT	AAGCCTACTC	GGCCACCGTC	AGAACTCCTCA	AGCAGCACCT	CCTCTTCATC	10260

40/56

Figure 7d (CONTINUED 3)

TTCCAGGCAG	TCATTAGGAC	TTTCTCTAAA	CAATTTGAAC	ATCACAGAGG	CTGTTAGCTC	10320
AGATATTTTG	CTAGATGATG	CTGGTGATGC	AACTGGACAT	AAAGATGGCC	GCAGTGTGAA	10380
AATTATAGTC	TCCATAAGCA	AGGGCTATGG	TCGAGCAAAG	GACCAAAAAT	CTCAGGCATA	10440
TTTGATAGGC	TCCATTGGTG	TTAGTGGAAA	AACCAAGTGG	GATGCTTAG	ATGGTGTAAT	10500
AAGACGTCTC	TTTAAGGAAT	ATGTATTCCG	AATTGATACA	TCCACTAGCC	TTGGTCTGAG	10560
CTCTGACTGC	ATTGCTAGCT	ACTGTATAGG	AGACTTAATT	AGATCCCAT	ACCTAGAAGT	10620
GCCTGAATTG	CTGCCTTGTG	GATACCTTGT	TGGAGATAAT	AACATCATCA	CTGTGAACCT	10680
CAAAGGGGTA	GAAGAAAATA	GTTTGGACAG	TTTTGTTTTT	GATACGCTGA	TTCCTAAACC	10740
AATTACCCAA	AGGTACTTTA	ACTTGTGAT	GGAGCATCAC	AGAATTATAC	TCTCAGGACC	10800
GAGTGGTACT	GGAAAGACCT	ATTGGGCAA	CAAACCTTGCT	GAATATGTAA	TAACCAAATC	10860
TGGAAGGAAA	AAAACAGAGG	ATGCAATTGC	CACTTTTAAT	GTGGACCACA	AGTCAAGTAA	10920
GGAAATGCAA	CAATATCTAG	CTAACCTGGC	TGAACAGTGC	AGTGCTGATA	ATAATGGAGT	10980
GGAGCTCCCA	GTTGTAATAA	TTCTTGATAA	TCCTCATCAT	GTGGGCTCTC	TGAGTGATAT	11040
CTTCAATGGT	TTTCTCAATT	GTAATACAA	CAATGTCCA	TATATTATTG	GAACAATGAA	11100
TCAGGGAGTT	TCTTCATCAC	CAAATCTAGA	GCTGCATCAC	AATTTCAAGT	GGGTATTATG	11160
TGCAAAATCAT	ACAGAACCAG	TGAAAGGCTT	TTTAGGCAGA	TATCTTCGAA	GAAAACCTCAT	11220
AGAGATAGAA	ATTGAAAGGA	ACATTCCGAA	TAATGACCTA	GTCAAAATTA	TAGATTGGAT	11280
TCCGAAGACG	TGGCATCATC	TCAACAGTTT	TTTGGAAACA	CACAGTTCTT	CTGACGTTAC	11340
CATTGGTCCC	CGACTATTCC	TTCTTGCCCC	CATGGATGTA	GAAGGTTCTA	GAGTATGGTT	11400
CATGGATCTC	TGGAACTATT	CTTTAGTACC	TTATATTCTG	GAGGCAGTGA	GAGAGGGTCT	11460
TCAGATGTAT	GGGAAACGCA	CACCATGGGA	AGATCCTTCA	AAGTGGGTGC	TTGACACATA	11520
TCCATGGAGC	TCAGCAACTC	TGCCTCAGGA	GAGCCCAGCC	TTACTTCAGC	TGCGACCAGA	11580
AGATGTTGGG	TATGAAAGCT	GCACATCCAC	TAAGGAAGCC	ACAACCTCAA	AGCACATTCC	11640
GCAAACCTGAC	ACAGAAGGAG	ATCCCCTGAT	GAATATGCTA	ATGAAACTCC	AAGAAGCAGC	11700
CAATTACTCA	AGCACACAAA	GCTGCGACAG	CGAAAGCACC	AGCCACCATG	AAGACATTTT	11760
GGATTTCATCT	CTTGAATCTA	CCCTCTAGAG	GGTGAAAGCC	GAAATCCAGC	ACACTGGCGG	11820
CCGTTACTAG	TGGATCGGCC	GC				11842

//

Legend: pGI3305 was obtained by inserting a 7148 bp XhoI/SacII fragment of the Hs-unc-53/3A1Ld22 clone in a XhoI/SacII opened pEGFPc3 vector (Clontech Inc.). This plasmid encodes an eGFP protein in fusion with the full length Hs-unc-53/3 (2363 AA). Arrows indicate the ORFs.

41/56

Figure 7e: Illustration of the AA sequence of GFP::Hs-unc-53/3 (insert of pGI3305)

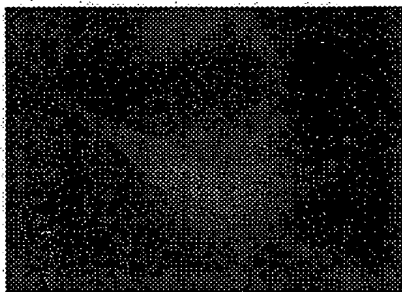
MVSKGEELFTGVVPIVELDGDVNGHKFSVSGEGEDATYGKLTTKFICTTGKLPVPWPTLVTTLTTYGV
QCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNIL
GHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSA
LSKDPNEKRDHMLLEFVTAAGITLGMDELYKYSDLEHMPVLGVASKLROPAVGSKPVHTALPIPNLGT
TGSOHCSSRPLELAETESSMLSCOLALKSTCEFGEKKPLOGKAKEKEDSKIYTDWANHYLAKSGHKRLI
KDLOODIADGVLLAEIIIOIIANEKVEDINGCPRSOSOMIENVDVCLSFLAARGVNVOGLSAEETRNGNL
KAILGLFFSLSRYKOOOHHOQOYYOSLVELQORVTHASPPSEASOAKTOODMOSSLAARYATOSNHSGI
ATSOKKPTRLPGPSRVPAAGSSSKVOGASNLNRRSOSFNSIDKNKPPNYANGNEKDSSKGPOSSSGVNG
NVOPPSTAGOPPASAIPSPSASKPWRSKSMNVKHSATSTMLTVKOSSTATSPTPSSDRLKPPVSEGVKT
APSGOKSMLEKFKLVNARTALRPPOPSSGSPDGGKDDDAFSESGEMEGFNSGLNSGGSTNSSPKVSPK
LAPPKAGSKNLSNKKSLLOPKEKEEKNRDKNKVCTEKPVKEEKDOVTEMAPKKTSKIASLIPKGSKTTA
AKKESLIPSSSGIPKPGSKVPTVKOTISPGSTASKESEKFRTTKGPSOSLSKPITMEKASASSCPAL
EGREAGOASPSGSCTMTVAOSSGOSTGNGAVOLPOOQOHSHPNTATVAPFIYRAHSENEGTALPSADSC
TSPTMDLSYSKTAKOCLEEISGEDPETRRMRTVKNIADLRONLEETMSSLRGTOISHSTLETTFDSTV
TTEVNGRTIPNLTSRPTPMTWRLGOACPRLOAGDAPSLGAGYPRSGTSRFIHTDPSRFMYTTPLRRAAV
SRLGNMSOIDMSEKASSDLMSSEVDVGGYMSDGDILGKSLRTDDINSGYMTDGGLNLYTRSLNRIPDT
ATSRDIIORGVHDVTVDADSWDDSSSVSSGLSDTLDNISTDDLNTTSSSVSSYSNITVPSRKNTOLRTDS
EKRSTTDETWDSPEELKKPEXDFDSHGDAGGKWKTVSSGLPEDPEKAGOKASLSVSOTGSWRRGMSAOG
GAPSROKAGTSALKTPGKTDDAKASEKGKAPLKGSSLORSPSDAGKSSGDEGKPPSGIGRSTATSSFG
FKKPSGVGSSAMITSSGATITSGSATLGKIPKSAAIGGKSNAGRKTSLDGSONODDVLHVSSKTTLOY
RSLPRPSKSSTSGIPGRGGHRSSTSSIDSNSVSSKSAGATTSKLREPTKIGSGRSSPVTVNOTDKEKEKV
AVSDSESVLSGSPKSSPTSASACGAOGLROPGSKYPDIASPTFRRLFGAKAGGKSASAPNTEGVKSSS
VMPSPSTTLAROGSLESPSSSGTGSMGSAGGLSGSSSPLFNKPSDLTTDVILSHSLASSPASVHSFTSG
GLVWAANMSSSSSAGSKDTPSYOSMTSLHTSSESIDLELSHHGSLSGLTTGTHEVOSLLMRTGSVRSTLS
ESMOLDRNTLPKKGLYTPSSROANOEEGKEWLRSHSTGLODTGNOSPLVSPSAMSSSAAGKYHFSNL
VSPTNLSOFNLPGPSMRRSNSIPAODSSFDLYDDSOLCGSATSLEERPRAISHSGSFRDSMEEVHGSSL
SLVSSTSSLYSTAEKAHSEQIHKLRRELVASOEKVATLTSOLSANAHLVAAFEKSLGNMTGRLOSLTM
TAEQKESELIELRETIEMLKAONSAAOAAIOGALNGDPHPKDLRIRROHSSESVSSINSATSHSSIGS
GNDADSKKKKKKNWNNSRGSELRSSFKOAFGKKKSTKPPSSHSDIEELTDSSLPASPKLPHNAGDCGSA
SMKPSOSASATCECTEAEAEITLQKSELREKELKLTDIRLEALSSAHHLDOIREAMNRMONEIEILKA
ENDRLKAETGNTAKPTRPPSESSSSSTSSSSRQSLGLSLNNLNITEAVSSDILLDDAGDATGHKDGRSV
KIIVSISKGYGRAKDQKSOAYLIGSIGVSGKTKWDVLDGVIRRLFEYVFRIDTSTSLGLSSDCIASYC
IGDLIRSHNLEVPELLPCGYLVGDNNIITVNLKVEENSLDSFVFDTLIPKPITORYFNLLMEHHRIIL
SGPSGTGKTYLANKLAEYVITKSGRKKTEDAIATFNVDHKSSKELOOYLANLAEOCSADNNGVELPVVI
ILDNLHHVGSLSDFNGFLNCKYNKCPYIIGTMNOGVSSSPNLELHHNFRWVLCANHTEPVKGLGRYL
RRKLEIEIERNIRNNDLVKIIDWIPKTWHHLNSFLETHSSSDVTIGPRLFLPCPMDVEGSRVWFMDLW
NYSLVPYILEAVREGLQMYGKRTPWEDPSKWVLDTYPWSSATLPQESPALLQLRPEDVGYESCTSTKEA
TTSKHIPOTDTEGDPLMNMLMKLOEAANYSSTQSCDSESTSHHEDILDSSLESTL

Legend: Single underlined AA sequence represents eGFP.
 Double underlined AA sequence represents full length Hs-unc-53/3.

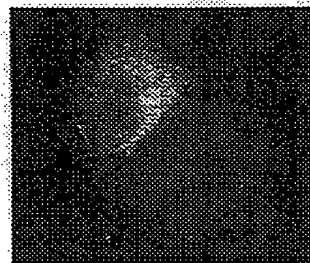
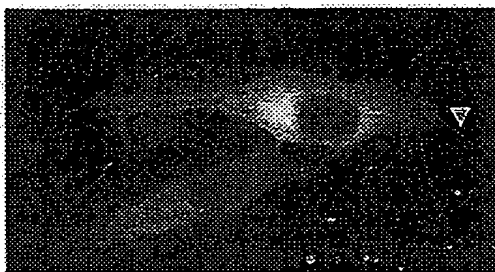
42/56

FIG. 8 Illustration of the filopodia and lamellipodia outgrowth of N4 mouse neuroblastoma cells transfected with pGI3303.

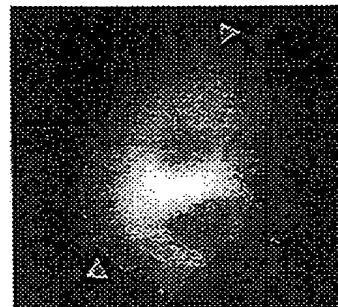
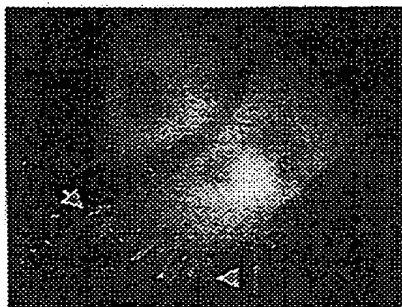
A:



B:



C:

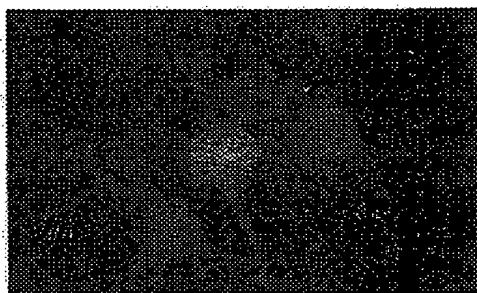


Legend: Fluorescence images of N4 cells transfected with pEGFP (A) compared to pGI3303 transfected cells (B and C). A: control (pEGFP) transfected cells. B: Illustration of filopodia outgrowth (arrowhead). C: Illustration of lamellipodia outgrowth (arrowhead). Notice the actin sheets at the edge of the cells.

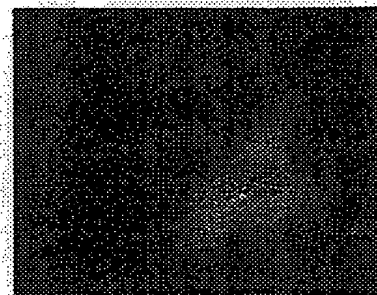
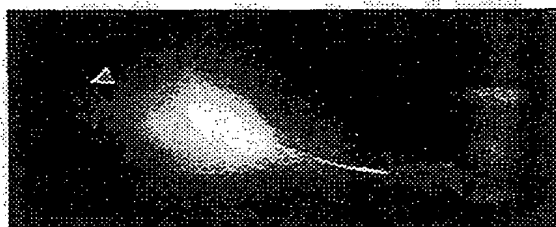
43/56

FIG. 9 Illustration of the co-localization of the GFP-Hs-unc-53/3 fusion protein with microtubules in N4 mouse neuroblastoma cells transfected with pGI3305

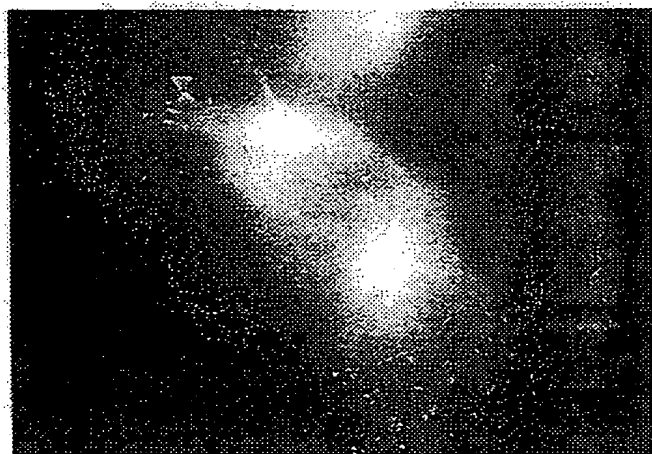
A:



B:



C:



Legend: Fluorescence images of N4 cells transfected with pEGFP (A) compared to pGI3305 transfected cells (B and C). A: control transfected cells. B: Illustration of co-localization of Hs-unc-53/3 with microtubuli. Notice the centrosome in the right picture (arrowhead) and enhanced filopodia outgrowth in the left picture (arrowhead). C: Illustration of the co-localization of Hs-unc-53/3 with (+)-end of microtubules (arrowhead).

44/56

Figure 11a: Illustration of the homology between Hs-unc-53/3 and a gene encoded (partially) by the *Drosophila melanogaster* BAC clone BACR48M05 (AC005719). Results of a TBLASTN search on the non-redundant database with Hs-unc-53/3 as query.

Query: Hs-unc-53/3 (direct) 2120aa Length 2119 from:1 to = 2119
 Sbjct: gb|AC005719|AC005719 *Drosophila melanogaster*, chromosome 2R, region
 38A5-38B4, BAC clone
 BACR48M05, complete sequence [*Drosophila melanogaster*]
 Length = 188357

Score = 64.0 bits (153), Expect = 4e-08
 Identities = 28/58 (48%), Positives = 41/58 (70%)

Query: 1 IYTDWANHYLAKSGHKRLIKDLQQDIADGVLLAEIIQIIAN EKVEDINGCPRSQSQMI 58
 IYTDWAN+YL ++ KR + DL D DG+LLAE+I- + + KV D+ P++Q QM+
 Sbjct: 84874 IYTDWANYLERAKSKRKVTDLSADCRDGLLLAEVIEAVTSFKVPDLVKKPKNQQQMV 84701

Score = 39.9 bits (91), Expect = 0.77
 Identities = 22/55 (40%), Positives = 34/55 (61%)

Query: 48 NGCPRSQSQMIENVDVCLSF LAARGVN-VQGLSAEEIRNGNLKAILGLFFSLSR YK 102
 N C Q +NV+ CL L ++ V ++ ++ +I G LKA+L LFF+LSR+K
 Sbjct: 55621 NSCSLFQ---FDNVNSCLHVLRSQSVGGLENITTNDICAGRLKAVLALFFALS RFK 55463

Score = 35.2 bits (79), Expect = 3.8
 Identities = 31/72 (43%), Positives = 45/72 (62%)

Query: 1266 LEERPRAISHSGSFRDSMEEVHGGSSLSLVSSSTSSLYSTAEKHAHSEIQIHKLRRELVASQE 1325
 L+ R + HS S VHGS SL+S SSLY AEE+ + +I +L+REL +++
 Sbjct: 13387 LKSRLMQLCHSVSV-----SVHGSAA SLLSGGSSLYGNABER-QAHEIRRLKRELQDARD 13226

Query: 1326 KVATLTSQLSAN 1337
 +V +L+SQLS N
 Sbjct: 13225 QVLSLSSQLSTN 13190

45/56

Figure 11b: Illustration of an ORF encoded by the *Drosophila melanogaster* BAC clone BACR48M05 (AC005719) as prediction by the computer program Fgene.

Output file for REVERSE STRAND of FGene
F469BE1C

```
length of sequence - 188357
number of predicted exons - 21
positions of predicted exons:
  4726 -      4757 w=   4.11 ORF:   4726 -      4755
  4816 -      4966 w=  20.57 ORF:   4817 -      4966
  5018 -      5318 w=  15.85 ORF:   5018 -      5317
  8693 -      8727 w=  14.75 ORF:   8695 -      8727
 38041 -     38265 w=   8.43 ORF:  38041 -     38265
 62411 -     62522 w=  10.60 ORF:  62411 -     62521
 74061 -     74692 w=  19.39 ORF:  74063 -     74692
103484 -    103654 w=  24.14 ORF: 103484 -    103654
132758 -    133134 w=  17.28 ORF: 132758 -    133132
153576 -    153706 w=  18.42 ORF: 153577 -    153705
154573 -    154681 w=  20.72 ORF: 154575 -    154679
154753 -    156246 w=  23.66 ORF: 154754 -    156244
160324 -    160375 w=   6.48 ORF: 160325 -    160375
161337 -    161421 w=   6.82 ORF: 161337 -    161420
171340 -    171756 w=  10.27 ORF: 171342 -    171755
171821 -    171965 w=  18.76 ORF: 171823 -    171963
172024 -    172326 w=  15.53 ORF: 172025 -    172324
174437 -    174810 w=   9.70 ORF: 174438 -    174809
175017 -    175168 w=  16.41 ORF: 175019 -    175168
179216 -    179267 w=   6.89 ORF: 179216 -    179266
187662 -    187678 w=   5.32 ORF: 187664 -    187678
```

Length of Coding region- 5367bp
Amino acid sequence - 1788aa

```
MDSGICYIKPEYLVTEADGGSAAANTENSNTNKRKREDGGEVEAGEKKKWDKKERKRGQN
KNRPVFKDERYSHLCHSLIDGTGGEPCLANCRYVHDLDAYLAAKGEDLGPECYVYTTKG
YCARGVSCRFAKAHTDEQGRNLKREDYDENAPPTTCNGVSSAASSTLHNASMOMNPLTNM
KNVLKLSEHELQHGKKSWHDMYKDSAWIFVAGFPYTLTEGDLVCVFSQYGEVVNINLIR
DSKTGKSKHSPLYRGEILFRIPELSQIPDPLCFLCNSIKLNSEVLNPNANFPMDIGIPNPY
TNEQLVNAKLEQQNLEKLFNELENTASMSNSQESKDTETTSTALVESSTSTNSASSAGSC
SLANPAQQSMKKKLTFLNLSPPFRSGKKSIDKNTSEQQRAISELVSTDHMLHLQQLLQQQR
KDQRSHTVPTESNYVLFNPGPVPSRHVQYKIRKPRPLSTHSDADSGFLSPCSPEEMRANP
AILVLQQCDVQGYMEIYTDWANYYLERAKSKRKVTDLSADCRDGLLLAEVIEAVTSFKV
PDLVKKPKNQQMFDNVNSCLHVLRSQSVGGLENITNDICAGRLKAVLALFFALSFRFKQ
QAKQTKSIGVGC GGVGSSSTLTGSGSVLGIGIGGLRTPGSSLNQDKNQEQEQEQEQEQ
QTPQQLAQSLNENEMVNRQIAPAYAKVNGGTAIPLPATVMVQRRCPDPKVRPLPPTPNH
TPSPGLGKSGSDFNTSRPNSPPTSNHTIQSLKSGNNNSLRPPSIKSGIPSPSPQTAPO
KHSMLDKLKLFNKEKQONAVNAASVASKTQIQSKRTSSSSSGFSSARSERSDSSLSLNDGH
GSQKPPSISVSSQKPPKTKQSKLLAAQQKKEQANKATKLDKKEKSPARSLNKEESGNE
SRSSMTGRTGKSSLVRAVGVEKNTPKTSSKSSLHSKSDSKSSLKAPQLLQSPSSGGLPK
PIAAIKGTSKPLSLGGAGHLPAAESQQNQQLLKRETS DISSNISQPPPAEPPISTHAHI
HQNQTPPPPYANSQPTSHISSHGFLSEPSTPOHSSGIYGSSRLPPPKSALSAPRKLEYN
AGPHILSSPTHHRQQLPRPLVNSAPNTPTASPNKFHTIPSKIVGTIYESKEEQLLPAPP
PASGGSSILPMRPLLRGYNSHVTLPTRGARGGHHPHQSYLDFCESDIGQGYCSDGDALRV
GSSPGGSRFHDIDNGYLSGSSGLNGPSSSAGGISPGKHFLSMMRARTQLPTTIEERQLI
YGASVPILTLLPDRKIYQNNVRQIKVDKLAAMAERWNMELGNGGAKMDGSPHHRPGSRNG
RDNWSKMPEPLNGQKVEKSDKSSPSRRSMGGGGSGSSSKQGPSSSSSRTKGVPSPFGYVK
RANGSIASAEQQNIAMMAAGGAGANGLPCGRTAHVSAVPRTASGRKVAGGTQTLPNDM
NKLPPNTQHRFSLTGTATQLSQSIRERLATGSHSLPKPGSDLHVFQHRISNRGGTRHD
```

46/56

Figure 11b (CONTINUED)

GSLSDTQTYAEVKPEYSSYAMWLKHSNTAGSRLSDGESVEQLQIGSPALTRHGHKMIHNR
SGGPGQMAGQMSGNESPYVQSPRMNRSNSIRSTKSEKMYPSMMSRAGEVEIEPYICLPVG
TNGVLTAAQMAAAQSQAAQGNPGVGVNVGGVAWSQPTSPPTPLTRGPFNTAAGASVLSP
THGTTSAAGLVGPGGGAGGGAMVGHRLTYPKKNDEVHGSAAASLLSGGSSLYGNAEERQAH
EIRRLKRELQDARDQVLSLSSQLSTNVSKKCPVVVFQMYTLRMARSRR*

47/56

Figure 11c: Illustration of a 'BLAST 2 sequences' search result with Hs-unc-53/3 as query and the Fgene predicted UNC53 homology ORF of Drosophila melanogaster BAC clone BACR48M05 as subject

Query: Hs-unc-53/3 (direct) 2120aa Length 2119 from:1 to = 2119
Subject: drosUNC53 (Fgene-prediction) Length 1788 from:1 to = 1788

Score = 106 bits (261), Expect = 2e-21
Identities = 190/840 (22%), Positives = 294/840 (34%), Gaps = 185/840 (22%)

```

Query: 1      IYTDWANHYLAKSGHKRLIKDLQQDIADGVLLAEIIQIIANEKVEDINGCPRSQSQMIE 60
IYTDWAN+YL ++ KR + DL D DG+LLAE+I+ + + KV D+ P++Q QM +N
Sbjct: 497    IYTDWANYLERAKSKRKVTDLSADCRDGLLLAEVIEAVTSFKVPDLVKKPKNQQMFDN 556

Query: 61     VDVCLSF LAARGV-NVQGLSAEEIRNGNLKAILGLFFSLSRK----- 102
V+ CL L ++ V ++ ++ +I G LKA+L LFF+LSR+K
Sbjct: 557    VNSCLHVLRSQSVGLENITTNDICAGRLKAVLALFFALS RFKQQAQTKSIGVCGGGV 616

Query: 103     XXXXXXXXXXXXSLVEL---QQRVTHASPPSEASQAKTQQDMQSSLAARYATQSNHSG--- 156
          S++ + R +S + +Q + QQ Q + QS +G
Sbjct: 617     GGSSSTLTGSGSVLGIGIGGLRTPGSSSLNQDKNQEQQQQQQQQQTPQQLAQSLNENEM 676

Query: 157     ----IATSQKK---PTRLPGPSRV-----PAAGSSSKVQGASNLNRRSQSFNS 197
          IA + K T +P P+ V P + + L + FN+
Sbjct: 677     VNRQIAPAYAKVNGGTAIPLPATVMVQRRCPDPKVRPLPPTPNHTPSIPGLGKSGSDFNT 736

Query: 198     IDKNKPPNYANGNEKDSSKGPQS--SSGVNGNVQPPSTAGQXXXXXXXXXXXXKPPWSKSM 256
          N PP S+ QS SG N +++PPS
Sbjct: 737     SRPNSPPT-----SNHTIQSLKSGMNNSLRPPSIKSGI----- 769

Query: 257     NVKHSATSTMLTVKQXXXXXXXXXXXXDLKPPVSEGVKTAPSGQKSMLEKFKLVNARTAL 316
          P +TAP + SML+K KL N
Sbjct: 770     -----PSPSSPQTAPQ-KHSMCLKLKFNKEKQQ 797

Query: 317     RXXXXXXXXXXXXXXXXXAFSESGEMEGFXXXXXXXXXXXXPKVSPKLAPPKAGSKNLS 376
          S SG +L PP S ++S
Sbjct: 798     NAVNAASVASKTQIQSKRTSSSSGFSS--ARSERSDSSLSLNDGHGSQKLP--SISVS 852

Query: 377     NKKSLLQPXXXXXXXXNRDKNKVCTEKPVKEEKDQVTEMAPKKTSKIASLIPKGSKTAAAK 436
++K QP ++K+ + KE+ ++ T++ K+ S SL + S + +
Sbjct: 853     SQKP--QP-----KTKQSKLLAAQKKEQANKATKLDKKEKSPARSLNKEESGNES--R 902

Query: 437     ESLXXXXXXXXXXXXXXXXTVKQITISPGSTASKSEKFRRTTKGSPSQSLSKPITMEKASAS 496
          S K T S +S S K SL P ++ S+
Sbjct: 903     SSTMGRTGKSSLVRVAVGVEKNTPKTSSKSSLHS-----KSDSKSSLKAPQLLQSPSSG 956

Query: 497     SCPAPLEGREAGQASPSGSCTMTVAQSSSQSTGNGAVQLP-----QQQHQSHPNTATVA- 550
          P P+ + P S G GA LP Q QQ T+ ++
Sbjct: 957     GLPKPIAAIKGTSKLP-----SLGGGAGHLPAAESQQNQQLLKRETSDISS 1002

Query: 551     -----PFIYRAHSENEG TALPSADCTSP TKMDLSYSKTAKQCLEEISGEGPETR 600
          P AH T P + + PT S+ ++ S +
Sbjct: 1003    NISQPPPAEPISTHAHIHQNTPPPPYYANSQPTSHISSHGFLSEPSTPQHSSGIYGSS 1062

Query: 601     RMRTVKNIADLRQNL EETMSSLRGTQISHSTLETTFDSTVTTEVNGRTI-PN-LTSRPTP 658
R+ K+ + LE + +H + V + N T PN + P+
Sbjct: 1063    RLPPFKSALSAPRKLEYNAGPHILSSPTHHQRQGLRPLVNSAPNTPPTASPKNFHTIPSK 1122

Query: 659     MTWRLGQACPR LQAGDAPSLGAGYPRSGTSRFIHTDPSRFMY----TTPLRRAAVSRLGN 714
+ + ++ + L A P SG S + P Y T P R A +
Sbjct: 1123    IVGTI-----YESKEEQLLPAPPPASGGSSILPMRPLLRGYNHSHVTLPTRGARGGHHHP 1176

Query: 715     MSQIDMSEKASSDLDMSSEVDVG-GYMSDGDIL--GKS---LRTDDINSGYMTDG--GLN 766
          S +D E D+G GY SDGD L G S R DI++GY+++G GLN
Sbjct: 1177    QSYLDFCES-----DIGQGYCSDGDALRVGSSPGGSRFHDIDNGYLSEGSGLN 1225

```

48/56

Figure 12: Illustration of an EST encoding a part of the Zebrafish-UNC-53/2 cDNA.

Query= hh2UNC53 (2340 letters)

Sbjct= emb|AI658309|AI658309 fc21d06.y1 Zebrafish WashU MPIMG EST Danio rerio cDNA 5' similar to TR:Q20427 Q20427 F45E10.1 mRNA sequence. Length = 445

Score = 277 bits (702), Expect = 4e-73
 Identities = 124/147 (84%), Positives = 136/147 (92%)
 Frame = +3

Query: 2121 LHHNFRWVLCANHTEPVKGFLGRFLRRKLMETEISGRVRNMELVKIIDWIPKVWHHLNRF 2180
 LHHNFRW+LCANHTEPVKGFLGRFLRRKL+ETEI+ RVRN ELVKII+WIP VWHHLNRF
 Sbjct: 3 LHHNFRWILCANHTEPVKGFLGRFLRRKLLETEINSRVRNGELVKIIEWIPSVWHHLNRF 182

Query: 2181 LEAHSSSDVTIGPRLFLSCPIDVDGSRVWFDTLWNYSIIPYLLEAVREGLQLYGRRAPWE 2240
 LE HSSSDVTIGPRLFLSCP+DV+GSRVWFDTLWNYSIIPY+LEAVREGLQ+YGR+A WE
 Sbjct: 183 LETHSSSDVTIGPRLFLSCPMDVEGSRVWFDTLWNYSIIPYMLEAVREGLQMYGRKASWE 362

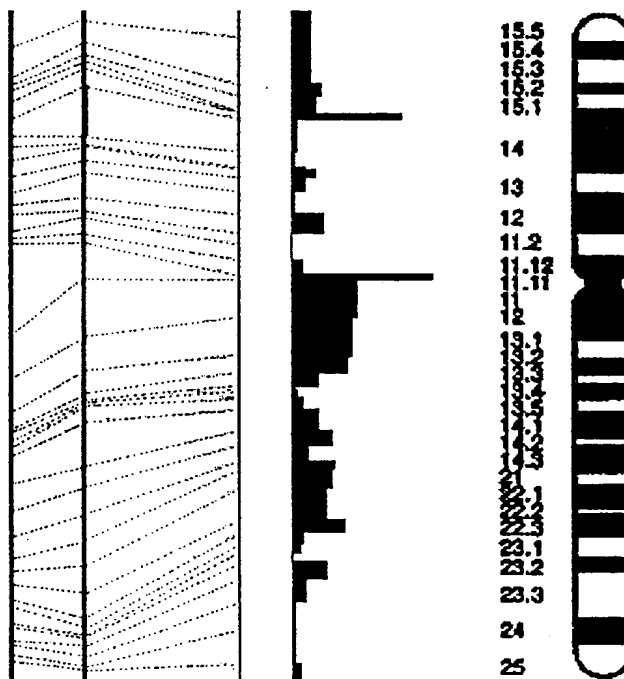
Query: 2241 DPAKWVMDTYPWAASPPQHEWPPLLQL 2267
 DPAKWVM++ ASPQHEW LL+L
 Sbjct: 363 DPAKWMEsLLCVASPPQHEWHSLLRL 443

49/56

Figure 13. Genemap98 results for Hs-Unc53/2

UniGene	Hs.13830		
RH Mapping Results			
SHGC-33456	G3 Map:		Chr.11
	Reference interval:		D11S921-D11S1359 (24.9-32.5 cM)
	Physical position:		911 cR10000 (F)
	RH details:		RHdb RH32790
	Typed by:		Stanford (see SHGC-33456)
Electronic PCR Results			
ESTs (from GenBank EST division)			
AA115015	zl04d10.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 491347 3'		
	STS	7 ... 134 bp:	SHGC-33456
AA918601	ol53e11.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1527212 3'		
	STS	16 ... 143 bp:	SHGC-33456
AJ248585	qh71f08.x1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:1850151 3', mRNA sequence [Homo sapiens]		
	STS	19 ... 146 bp:	SHGC-33456
T71262	yd35b09.s1 Homo sapiens cDNA clone 110201 3'.		
	STS	9 ... 136 bp:	SHGC-33456

RH Map Genetic Gene Cytogenetic
GB4 G3 Map Density Ideogram

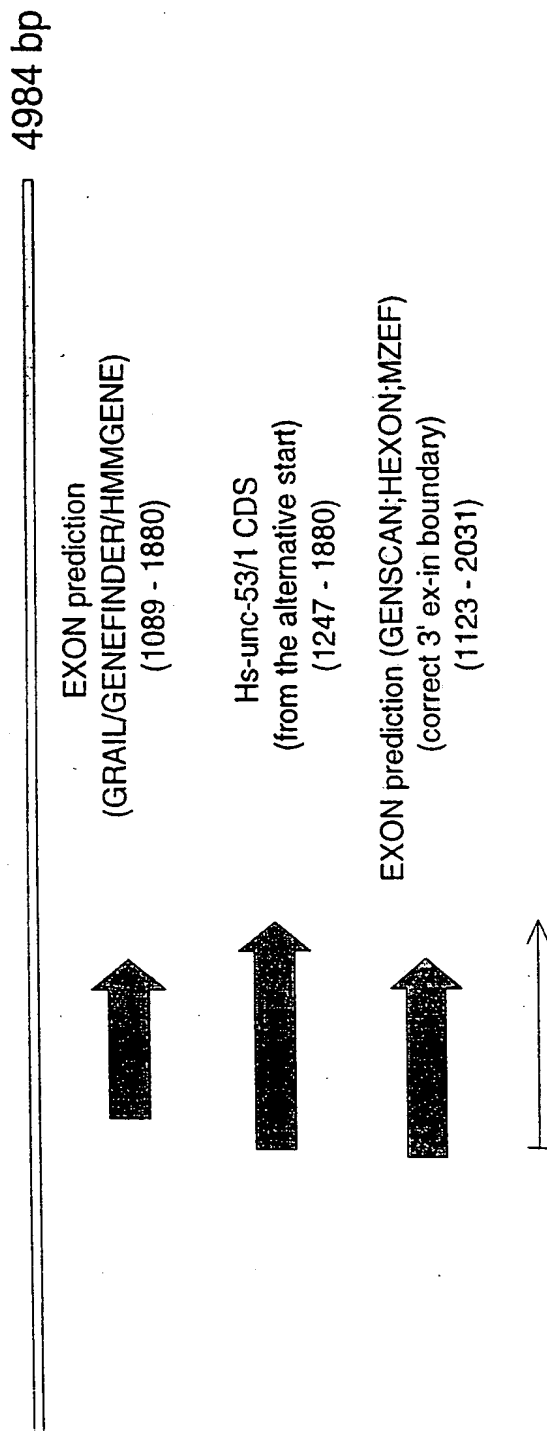


The thick line on the G3 map indicates the position of SHGC-33456 See also: equivalent interval on GB4 map

About This Interval	
Top of interval:	D11S921 (24.9 cM)
Bottom of interval:	D11S1359 (32.5 cM)
Genetic size of bin:	8 cM
Physical size of bin:	430 cR10000

50/56

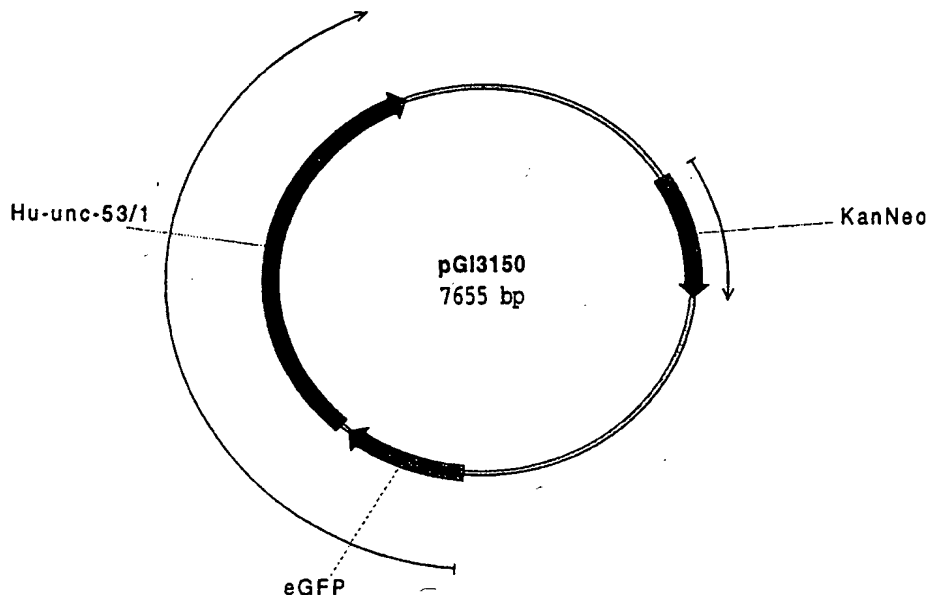
Figure 14. Prediction of a 5' exon of Hs-unc-53/1*



(*) numbers refer to figure 1g.

51/56

Figure 15: Illustration of the nucleotide sequence of pGI3150 and amino acid sequence of the eGFP fusion with a C-terminal fragment of Hs-Unc-53/1.



```

ID      pGI3150                                circular DNA; 7655 BP
DE      from coiled coil I till end
FT      CDS      1225..2019
FT                               /vntifkey="4"
FT                               /label=KanNeo
FT      CDS      3942..4658
FT                               /vntifkey="4"
FT                               /label=eGFP
FT      CDS      4719..7214
FT                               /vntifkey="4"
FT                               /label=Hu-unc-53/1
SQ      SEQUENCE      7655 BP;

```

```

CTAGATAACT GATCATAATC AGCCATACCA CATTGTGAGA GGTTTTACTT GCTTTAAAAA      60
ACCTCCCACA CCTCCCCCTG AACCTGAAAC ATAAATGAA TGCAATTGTT GTTGTTAACT      120
TGTTTATTGC AGCTTATAAT GGTTACAAAT AAAGCAATAG CATCACAAAT TTCACAAATA      180
AAGCATTTTT TTTACTGCAT TCTAGTTGTG GTTTGTCCAA ACTCATCAAT GTATCTTAAC      240
GCGTAAATTG TAAGCGTTAA TATTTTGTGA AAATTCGCGT TAAATTTTGT TTTAAATCAGC      300
TCATTTTTTA ACCAATAGGC CGAAATCGGC AAAATCCCTT ATAAATCAAA AGAATAGACC      360
GAGATAGGGT TGAGTGTGTG TCCAGTTTGG AACAGAGTC CACTATTAAA GAACGTGGAC      420
TCCAACGTCA AAGGGCGAAA AACCGTCTAT CAGGGCGATG GCCCACTACG TGAACCATCA      480
CCCTAATCAA GTTTTTTGGG GTCGAGGTGC CGTAAAGCAC TAAATCGGAA CCCTAAAGGG      540
AGCCCCGAT TTAGAGCTTG ACGGGGAAAG CCGGCGAAGC TGGCGAGAAA GGAAGGGAAG      600
AAAGCGAAAG GAGCGGGCGC TAGGGCGCTG GCAAGTGTAG CGGTCACGCT GCGCGTAACC      660
ACCACACCCG CCGCGCTTAA TGCGCCGCTA CAGGGCGCGT CAGGTGGCAC TTTTCGGGGA      720
AATGTGCGCG GAACCCCTAT TTGTTTATTT TTCTAAATAC ATTCAAATAT GTATCCGCTC      780
ATGAGACAAAT AACCTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG TCCTGAGCGC      840
GAAAGAACCA GCTGTGGAAT GTGTGTCAGT TAGGGTGTGG AAAGTCCCCA GGCTCCCCAG      900
CAGGCAGAAAG TATGCAAAGC ATGCATCTCA ATTAGTCAGC AACCAGGTGT GGAAGTCCCC      960
CAGGCTCCCC AGCAGGCAGA AGTATGCAAA GCATGCATCT CAATTAGTCA GCAACCATAG      1020
TCCCGCCCCCT AACTCCGCCC ATCCCGCCCC TAACTCCGCC CAGTTCCGCC CATTCTCCGC      1080
CCCATTGGCTG ACTAATTTTT TTTATTTATG CAGAGGCCGA GGCCGCCTCG GCCTCTGAGC      1140
TATTCAGAAA GTAGTGAGGA GGCTTTTGTG GAGGCCTAGG CTTTGTCAAA GATCGATCAA      1200
GAGACAGGAT GAGGATCGTT TCGCATGATT GAACAAGATG GATTGCACGC AGGTTCTCCG      1260
GCCGCTTGGG TGGAGAGGCT ATTCCGGCTAT GACTGGGCAC AACAGACAAT CGGCTGCTCT      1320
GATGCGCGCG TGTCCGGCT CAGCAGCGAG GGGCGCCCGG TTCTTTTGT CAAGACCGAC      1380
CTGTCCGGTG CCCTGAATGA ACTGCAAGAC GAGGCAGCGC GGCTATCGTG GCTGGCCACG      1440
ACGGGCGTTC CTTGCGCAGC TGTGCTCGAC GTTGTCAGT AAGCGGGAAG GGACTGGCTG      1500

```

52/56

Figure 15 (CONTINUED 1)

CTATTGGGCG	AAGTGCCGGG	GCAGGATCTC	CTGTCACTCT	ACCTTGCTCC	TGCCGAGAAA	1560
GTATCCATCA	TGGCTGATGC	AATGCGGGG	CTGCATACGC	TTGATCCGGC	TACCTGCCCA	1620
TTGACCAACC	AAGCGAAACA	TCGCATCGAG	CGAGCACGTA	CTCGGATGGA	AGCCGGTCTT	1680
GTCGATCAGG	ATGATCTGGA	CGAAGAGCAT	CAGGGGCTCG	CGCCAGCCGA	ACTGTTCCGC	1740
AGGCTCAAGG	CGAGCATGCC	CGACGGCGAG	GATCTCGTCG	TGACCCATGG	CGATGCCTGC	1800
TTGCCGAATA	TCATGGTGGA	AAATGGCCGC	TTTTCTGGAT	TCATCGACTG	TGGCCGGCTG	1860
GGTGTGGCGG	ACCGCTATCA	GGACATAGCG	TTGGCTACCC	GTGATATTGC	TGAAGAGCTT	1920
GGCGGGCAAT	GGGCTGACCG	CTTCCTCGTG	CTTTACGGTA	TCGCCGCTCC	CGATTCCGAG	1980
CGCATCGCCT	TCTATCGCCT	TCTTGACGAG	TTCTTCTGAG	CGGGACTCTG	GGGTTCGAAA	2040
TGACCGACCA	AGCGACGCCC	AACCTGCCAT	CACGAGATTT	CGATTCCACC	GCCGCTTCT	2100
ATGAAAGGTT	GGGCTTCGGA	ATCGTTTTC	GGGACGCCGG	CTGGATGATC	CTCCAGCGCG	2160
GGGATCTCAT	GCTGGAGTTC	TTGCCCCACC	CTAGGGGGAG	GCTAACTGAA	ACACGGAAGG	2220
AGACAATACC	GGAAGGAACC	CGCGTATGA	CGGCAATAAA	AAGACAGAAT	AAAACGACG	2280
GTGTGGGTC	GTTTGTTCAT	AAACGCGGGG	TTCGGTCCCA	GGGCTGGCAC	TCTGTCGATA	2340
CCCCACCGAG	ACCCCAATTG	GGCCAATACG	CCCGCGTTTC	TTCTTTTCC	CCACCCACC	2400
CCCCAAGTTC	GGGTGAAGCG	CCAGGGCTCG	CAGCCAAAGT	CGGGCGGGCA	GGCCCTGCCA	2460
TAGCCTCAGG	TTACTCATAT	ATACTTTAGA	TTGATTAAAA	ACTTCATTTT	TAATTTAAAA	2520
GGATCTAGGT	GAAGATCCTT	TTTGATAATC	TCATGACCAA	AATCCCTTAA	CGTGAGTTTT	2580
CGTTCCACTG	AGCGTCAGAC	CCCGTAGAAA	AGATCAAAGG	ATCTTCTTGA	GATCCTTTTT	2640
TTCTGCGCGT	AATCTGCTGC	TTGCAAAACA	AAAAACCACC	GCTACCAGCG	GTGGTTTGT	2700
TGCCGGATCA	AGAGCTACCA	ACTCTTTTTC	CGAAGGTAAC	TGGCTTCAGC	AGAGCGCAGA	2760
TACCAATATC	TGTCCTTCTA	GTGTAGCCGT	AGTTAGGCCA	CCACTTCAAG	AACTCTGTAG	2820
CACCGCCTAC	ATACCTCGCT	CTGCTAATCC	TGTTACCAGT	GGCTGCTGCC	AGTGGCGATA	2880
AGTCGTGTCT	TACCGGGTTG	GACTCAAGAC	GATAGTTACC	GGATAAGGGC	CAGCGGTCCG	2940
GCTGAACGGG	GGGTTCTGTC	ACACAGCCCA	GCTTGGAGCG	AACGACCTAC	ACCGAACTGA	3000
GATACCTACA	GCGTGAGCTA	TGAGAAAGCG	CCACGCTTCC	CGAAGGGAGA	AAGGCGGACA	3060
GGTATCCGGT	AAGCGGCAGG	GTGGAACAG	GAGAGCGCAC	GAGGGAGCTT	CCAGGGGGAA	3120
ACGCCCTGGT	TCTTTATAGT	CCTGTCGGGT	TTGCCCCACT	CTGACTTGAG	CGTCGATTTT	3180
TGTGATGCTC	GTCAGGGGGG	CGGAGCCTAT	GGAAAAACGC	CAGCAACGCG	GCCTTTTAC	3240
GGTTCTCGGC	CTTTTGCTGG	CCTTTTGCTC	ACATGTTCTT	TCCTGCGTTA	TCCCTGATT	3300
CTGTGGATAA	CCGTATTACC	GCCATGCATT	AGTTATTAAT	AGTAATCAAT	TACGGGGTCA	3360
TTAGTTTCATA	GCCCATATAT	GGAGTTCCGC	GTTACATAAC	TTACGGTAAA	TGGCCCGCCT	3420
GGCTGACCGC	CCAACGACCC	CCGCCCATTG	ACGTCAATAA	TGACGTATGT	TCCCATAGTA	3480
ACGCCAATAG	GGACTTTCCA	TTGACGTCAA	TGGGTGGAGT	ATTTACGGTA	AACTGCCACC	3540
TTGGCAGTAC	ATCAAGTGTA	TCATATGCCA	AGTACGCCCC	CTATTGACGT	CAATGACGGT	3600
AAATGGCCCG	CCTGGCATT	TGCCAGTAC	ATGACCTTAT	GGGACTTTCC	TACTTGGCAG	3660
TACATCTACG	TATTAGTCAT	CGCTATTACC	ATGGTGATGC	GGTTTTGGCA	GTACATCAAT	3720
GGCGCTGGAT	AGCGGTTTGA	CTCACGGGGA	TTTCCAAAGT	TCCACCCCAT	TGACGTCAT	3780
GGGAGTTTGT	TTTGGCACCA	AAATCAACGG	GACTTTCCAA	AATGTGCTAA	CAACTCCGCC	3840
CCATTGACGC	AAATGGGCGG	TAGGCGTGTA	CGGTGGGAGG	TCTATATAAG	CAGAGCTGGT	3900
TTAGTGAACC	GTGAGTCCG	CTAGCGCTAC	CGGTGCGCAC	CATGGTGAGC	AAGGCGGAGG	3960
AGCTGTTTAC	CGGGGTGGTG	CCCATCCTGG	TCGAGCTGGA	CGGCGACGTA	AACGGCCACA	4020
AGTTACAGCGT	GTCCGGCGAG	GGCGAGGGCG	ATGCCACCTA	CGGCAAGCTG	ACCCTGAAGT	4080
TCATCTGCAC	CACCGGCAAG	CTGCCCCGTC	CCTGGCCCCA	CCTCGTGACC	ACCCTGACCT	4140
ACGGCGTGCA	GTGCTTCAGC	CGCTACCCCG	ACCACATGAA	GCAGCACGAC	TTCTTCAAGT	4200
CCGCCATGCC	CGAAGGCTAC	GTCCAGGAGC	GCACCATCTT	CTTCAAGGAC	GACGGCAACT	4260
ACAAGACCCG	CGCCGAGGTG	AAGTTCGAGG	GCGACACCCT	GGTGAACCG	ATCGAGCTGA	4320
AGGGCATCGA	CTTCAAGGAG	GACGGCAACA	TCCTGGGGCA	CAAGCTGGAG	TACAACCTACA	4380
ACAGCCACAA	CGTCTATATC	ATGGCCGACA	AGCAGAAGAA	CGGCATCAAG	GTGAACTTCA	4440
AGATCCGCCA	CAACATCGAG	GACGGCAGCG	TGCAGCTCGC	CGACCACTAC	CAGCAGAACA	4500
CCCCCATCGG	CGACGGCCCC	GTGCTGCTGC	CCGACAACCA	CTACCTGAGC	ACCCAGTCCG	4560
CCCTGAGCAA	AGACCCCAAC	GAGAAGCGCG	ATCACATGGT	CCTGTGGAG	TTCTGTGACC	4620
CCGCCGGGAT	CACTCTCGGC	ATGGACGAGC	TGTACAAGTC	CGGACTCAGA	TCTCGAGCTC	4680
AAGCTTCGAA	TTCTGCAGTC	GACGGTACCG	CGGGCCCCGG	ATCCTTCCGA	GACCCACCGG	4740
ACGATGTTCA	CGGCTCAGTG	CTGTCCCTGG	CCTCCAGTGC	CTCCTCCACC	TACTCCTCAG	4800
CTGAGGAGAG	GATGCAATCT	GAGCAAAATCC	GGAAGCTTCG	TAGGGAACTG	GAATCATCCC	4860
AGGAAAAAGT	GGCCACCTTG	ACGTCTCAGC	TTTCTGCCAA	TGCTAATCTG	GTGGCTGCTT	4920
TTGAGCAGAG	CCTGGTGAAT	ATGACATCCC	GCCTGCGACA	CCTGGCAGAG	ACGGCCGAGG	4980
AGAAGGACAC	TGAGCTGCTG	GATTTGGCAG	AAACCATAGA	CTTTCTGAAG	AAAAAGAACT	5040
CTGAGGCCCA	GGCAGTCATT	CAGGAGCCCC	TTAATGCCTC	AGAAACCCCA	CCCAAAGAAC	5100
TTCGGATCAA	GAGACAAAAC	TCCTCAGATA	GCATCTCAAG	CCTCAACAGC	ATCACTAGCC	5160
ATTCCAGCAT	CGGCAGCAGC	AAGGATGCTG	ATGCGAAAAA	GAAGAAAAAA	AAGAGTTGGG	5220
TCTATGAGCT	TGGAAGTTCC	TTCAACAAAG	CGTTCAAGT	AAAAAGGGG	CCCAAGTCAG	5280
CTTCCTCATA	CTCGGATATA	GAGGAGATTG	CTACACCCGA	CTCTTCAGCC	CCCTCATCCC	5340
CCAAACTACA	GCATGGTTCC	ACAGAGACTG	CTTCACCCCT	CATCAAGTCC	TCCACCTTGT	5400
CCTCCGTGGG	CACTGATGTC	ACCGAGGGCC	CTGCTCACCC	AGCCCCCCAC	ACTAGGCTGT	5460
TCCATGCAAA	TGAGGAGGAG	GAGCCAGAGA	AGAAGGAGGT	ATCGGAGCTG	CGCTCTGAGC	5520
TATGGGAGAA	GGAAATGAAG	CTTACAGACA	TCCGCTTGGA	GGCCCTCAAC	TCTGCCCAAC	5580
AACTGGATCA	GCTTCGGGAG	ACCATGCACA	ACATGCAGTT	GGAGGTGGAC	CTGCTGAAAG	5640
CAGAGAAATGA	CCGACTGAAG	GAGCCCCAG	GCCCTCATC	AGGCTCCACT	CCAGGGCAGG	5700
TCCCTGGATC	ATCTGCATTA	TCTTCCCCAC	GCCGCTCCCT	AGGCCTGGCA	CTCACCATT	5760
CCTTCGGCCC	CAGTCTTGCA	GACACAGACC	TGTCACCCAT	GGATGGCATC	AGTACTGTGT	5820

53/56

Figure 15 (CONTINUED 2)

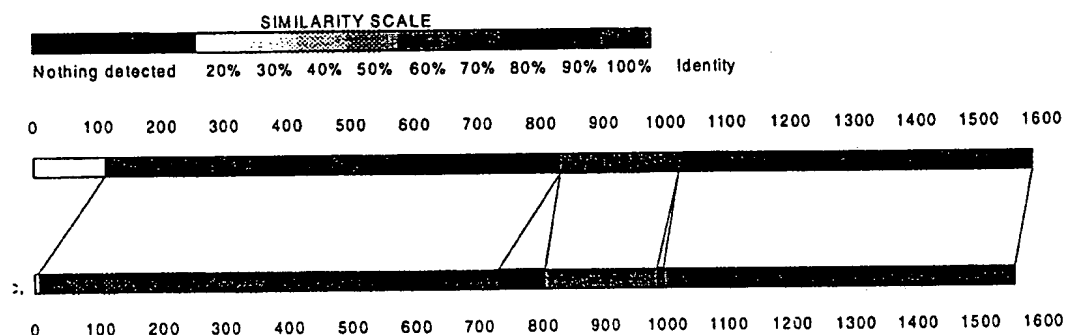
GTCCAAAGGA	GGAAGTGACC	CTCCGGGTGG	TGGTGAGGAT	GCCCCGCAG	CACATCATCA	5880
AAGGGGACTT	GAAGCAGCAG	GAATTCCTCC	TGGGCTGTAG	CAAGGTCAGT	GGAAAAGTTG	5940
ACTGGAAGAT	GCTGGATGAA	GCTGTTTTCC	AAGTGTTCAA	GGACTATATT	TCTAAAATGG	6000
ACCCAGCCTC	TACCCTGGGA	CTAAGCACTG	AGTCCATCCA	TGGCTACAGC	ATCAGCCACG	6060
TGAAACGAGT	GTTGGATGCA	GAGCCCCCGG	AGATGCCTCC	TTGCCGTCGA	GGTGTCAATA	6120
ACATATCAGT	CTCCCTCAAA	GGTCTGAAGG	AGAAATGCGT	CGACAGCCTG	GTGTTTCGAGA	6180
CGCTGATCCC	CAAGCCGATG	ATGCAGCACT	ACATAAGCCT	CCTGCTGAAG	CACCGGCGCC	6240
TCGTCTCTCT	GGGCCCCCAGC	GGCACGGGCA	AGACCTACCT	GACCAATCGC	TTGGCCGAGT	6300
ACCTGGTGGA	GCGCTCTGGC	CGTGAGGTCA	CAGAGGGCAT	CGTCAGCACC	TTCAACATGC	6360
ACCAGCAGTC	TTGCAAGGAT	CTGCAACTGT	ATCTTTCCAA	CCTAGCCAAC	CAGATAGACC	6420
GGGAAACAGG	AATTGGGGAT	GTGCCCTTGG	TGATTCTATT	GGATGACCTG	AGTGAAGCAG	6480
GCTCCATCAG	TGAGTTGGTC	AATGGGGCCC	TCACCTGCAA	GTATCATAAA	TGTCCTTATA	6540
TTATAGGTAC	CACCAATCAG	CCTGTAAAAA	TGACACCCAA	CCATGGCTTG	CACCTTGAGCT	6600
TCAGGATGTT	GACCTTCTCC	AACAACGTGG	AGCCAGCCAA	TGGCTTCTCT	GTTTCGTTACC	6660
TGAGGAGGAA	GCTGGTAGAG	TCAGACAGCG	ACATCAATGC	CAACAAGGAA	GAGCTGCTTC	6720
GGGTGCTCGA	CTGGGTACCC	AAGCTGTGGT	ATCATCTCCA	CACCTTCCTT	GAGAAGCACA	6780
GCACCTCAGA	CTTCCTCATC	GGCCCTTGCT	TCTTTCTGTC	GTGTCCCAT	GGCATTGAGG	6840
ACTTCCGGAC	CTGGTTCAAT	GACCTGTGGA	ACAACCTCTAT	CATTCCCTAT	CTACAGGAAG	6900
GAGCCAAGGA	TGGGATAAAG	GTCCATGGAC	AGAAAGCTGC	TTGGGAGGAC	CCAGTGGAAT	6960
GGGTCCGGGA	CACACTTCCC	TGGCCATCAG	CCCAACAAGA	CCAATCAAAG	CTGTACCACC	7020
TGCCCCCACC	CACCGTGGGC	CCTCACAGCA	TTGCCTCACC	TCCCGAGGAT	AGGACAGTCA	7080
AAGACAGCAC	CCCAAGTTCT	CTGGACTCAG	ATCCTCTGAT	GGCCATGCTG	CTGAAACTTC	7140
AAGAAGCTGC	CAACTACATT	GAGTCTCCAG	ATCGAGAAAC	CATCCTGGAC	CCCAACCTTC	7200
AGGCAACACT	TTAAGGGTTC	GGCAATCACT	GTCACCCCCG	GACAGCAGAA	CGCTGGCATC	7260
AGCTATCTTA	GCTCCTCTCT	TCCCCTCTCC	TCTTTACAGAG	CACTGGCTCT	CCAGCCCCAG	7320
GAGGAGAACA	GGAGGGAGGA	GGAGATGAAA	GAGGAGGGAC	AGGTTCCTGG	TGCTGTACCT	7380
TTGAGAACTT	CCTAGGAAGG	AATGGTGGGG	TGGCGTTTGG	GAACCTGTGC	CCCCTAAACA	7440
CATTTACTGG	CCTCCTCTAA	TGACTTTGGG	GAAAAGATGA	TTCTGGGTCT	TTCCCTTGAC	7500
TTCTTGTTTC	AATTACAAAC	TCCTGGGCTT	TCTGGGGAGG	GGTTCAGAAA	ACATCAAAAC	7560
ACTGCAGCAG	TTCCCCGGAA	TTCAAGCTTG	ACTTAACCAG	GCTGAACCTG	CTCAAAAGAA	7620
GCCGAATTCC	AGCACACTGG	CGGCCGTTAC	TAGTT			7655

//

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPPTLVTTLTYGVCQ
 FSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKL
 EYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPN
 EKRDMVLLEFVTAAGITLGMDELYKSGLSRAQASNSAVDGTAGPGSFRDPTDDVHGSVLSLASSASTY
 SSAEERMQSEQIRKLRRELESSQEKVATLTSQLSANANLVAAFEQSLVNMTSRLRHLAETAEEKDTELLDL
 RETIDFLKKKNSEAQAVIQALNASETTPKELRIKRONSSDSISSLNSITSHSSIGSSKDADAKKKKKKSW
 VYELRSSFNKAFSIKKGPKSASSYSIDIEEATPDSSAPSSPKLQHGSTETASPSIKSSTLSSVGTDVTEGP
 AHPAPHTRLRFHANESEPEKKEVSELRLSELWEKEMKLTDIRLEALNSAHQLDQLRETMHNMQLEVDLLKAE
 NDRILKVAPGPGSSGSTPGQVPGSSALSSPRRSILGLALTHSFGPSLADTDLSMPDGI STCGPKKEEVLTRVVVR
 MPPQHIKGLDKQOEFFLGCSKVSQKVDWKMLEDVAFQVQFDYISKMDPASTLGLSTESIHGYSISHVKRV
 LDAEPPPEMPPCRRGVNINISVSLKGLKEKCVDSLVEFTLIPKPMQHYISLLLKHRLVLSGPSGTGKTYLT
 NRLAEYLVERSGREGVTEGIVSTFNMHQQSCDLQLYLSNLANQIDRETGIGDVPLVILLDDLLSEAGSISEL
 VNGALTCKYHKCPYIIGTTNQPVKMTPNHGLHLSFRMLTFSNNVEPANGFLVRYLRRKLVESDSIDINANKE
 ELLRVLDWVPLWYHLHTFLEKHSTSDFLIGPCFFLSCPIGIEDFRTWFDLWNNISII PYLQEGAKDGIKV
 HGQKAWEDEPVEWVRDTPWPSAQQDQSKLYHLPPTVGPHSIASPPEDRTVKDSTPSSLDSDPLMAMLLK
 LQEAANYIESPDRETILDPNLQATL

54/56

Figure 16: EST Clone yk480b6 contains a splice variant of Ce-UNC-53



Results of SIM with:

Sequence 1: Ce-unc-53, (1583 residues) Sequence 2: yk480b06rc, (1556 residues)

Ce-UNC-53	110	LSTYKQKLRQLKKDQKKLEQLPTSIMP	PAVSKLPSRVATSATASATNPNSNFPQMSTSR
yk480b06rc	5	IQEFGTRLRQLKKDQKKLEQLPTSIMP	PAVSKLPSRVATSATASATNPNSNFPQMSTSR
Ce-UNC-53	170	LQTPQSRISKIDSSKIGIKPKTSGLKPPSSSTTSSNNTNSFRPSSRSSGNNVNGSTISTS	
yk480b06rc	65	LQTPQSRISKIDSSKIGIKPKTSGLKPPSSSTTSSNNTNSFRPSSRSSGNNVNGSTISTS	
Ce-UNC-53	230	AKSLESSSTYSSISNLNRPTSQLQKPSRPQTQLVRVATTTKIGSSKLAAPKAVSTPKLAS	
yk480b06rc	125	AKSLESSSTYSSISNLNRPTSQLQKPSRPQTQLVRVATTTKIGSSKLAAPKAVSTPKLAS	
Ce-UNC-53	290	VKTIGAKQEPDNSGGGGGMLKLKLFSSKNPSSSSNSPQPTRKAAAVPQQQTLSKIAAPV	
yk480b06rc	185	VKTIGAKQEPDNSGGGGGMLKLKLFSSKNPSSSSNSPQPTRKAAAVPQQQTLSKIAAPV	
Ce-UNC-53	350	KSGLKPPTSKLGSATSMKLCPTPKVSYRKTDAPII SQQDSKRC SKSSEESGYAGFNST	
yk480b06rc	245	KSGLKPPTSKLGSATSMKLCPTPKVSYRKTDAPII SQQDSKRC SKSSEESGYAGFNST	
Ce-UNC-53	410	PTSSSTEGSLMHSTSSKSSTSDKSPSSDDLTLNASIVTAIRQPIAATPVSPNI INKPV	
yk480b06rc	305	PTSSSTEGSLMHSTSSKSSTSDKSPSSDDLTLNASIVTAIRQPIAATPVSPNI INKPV	
Ce-UNC-53	470	EEKPTLAVKGVKSTAKKDPPAVPPRDTQPTIGVVSPIMAHKKLTNDPVI SEKPEPEKLQ	
yk480b06rc	365	EEKPTLAVKGVKSTAKKDPPAVPPRDTQPTIGVVSPIMAHKKLTNDPVI SEKPEPEKLQ	
Ce-UNC-53	530	SMSIDTTDVPPLPPLKSVVPLKMTSIRQPPTYDVLLKQGKITSPVKSPGYEQSSASEDSI	
yk480b06rc	425	SMSIDTTDVPPLPPLKSVVPLKMTSIRQPPTYDVLLKQGKITSPVKSPGYEQSSASEDSI	
Ce-UNC-53	590	VAHASAQVTPPTKTSGNHSLERRMGKNKTSSESSGYTSDAGVAMCAKMKREKLKEYDDMTRR	
yk480b06rc	485	VAHASAQVTPPTKTSGNHSLERRMGKNKTSSESSGYTSDAGVAMCAKMKREKLKEYDDMTRR	
Ce-UNC-53	650	AQNGYPDNFEDSSSLSSGISDNNELDDISTDDL SGVDMATVASKHSDYSHFVRHPTSSSS	
yk480b06rc	545	AQNGYPDNFEDSSSLSSGISDNNELDDISTDDL SGVDMATVASKHSDYSHFVRHPTSSSS	
Ce-UNC-53	710	KPRVPSRSSTSVDSRSRAEQENVYKLLSQCRTSQRGAAATSTFGQHSLSRSPGYSSYSPHL	
yk480b06rc	605	KPRVPSRSSTSVDSRSRAEQENVYKLLSQCRTSQRGAAATSTFGQHSLSRSPGYSSYSPHL	
Ce-UNC-53	770	SVSADKDTMSMHSQTSRRPSSQKPSYSGQFHS LDRKCHLQFTSTEHRMAALLSPRRVPN	
yk480b06rc	665	SVSADKDTMSMHSQTSRRPSSQKPSYSGQFHS LDRKCHLQFTSTEHRMAALLSPRRVPN	

55/56

Figure 16 (CONTINUED)

Ce-UNC-53	830	SMSKYDSS-----
yk480b06rc	725	SMSKYDSSAAALNASGMSRSMILLESLSRPPRRHQSPADSCIITASPSAPRRSHSPRGF

Ce-UNC-53	838	-----GSYSARSRGGSSTGIYGETFQLHRLSDEKSPAHSKSEMGS
yk480b06rc	785	TARIPLSLASSFVHVNNNNGSYSARSRGGSSTGIYGETFQLHRLSDEKSPAHSKSEMGS

Ce-UNC-53	879	QLSLASTTAYGSLNEKEYEHAIRDMARDLECYKNTVDSLTKKQENYGALFDLFEQKLRKLT
yk480b06rc	845	QLSLASTTAYGSLNEKEYEHAIRDMARDLECYKNTVDSLTKKQENYGALFDLFEQKLRKLT

Ce-UNC-53	939	QHIDRSNLKPEEAIRFRQDIAHLRDISNHLASNSAHANEGAGELLRQPSLESVASHRSSM
yk480b06rc	905	QHIDRSNLKPEEAIRFRQDIAHLRDISNHLASNSAHANEGAGELLRQPSLESVASHRSSM

Ce-UNC-53	999	SSSSKSSKQEKISLSSFGKNKKS-----IRSSLKFTKKKNKNYDEAHMPISIGSQG
yk480b06rc	965	SSSSKSSKQEKISLSSFGKNKKSALSVDSQIRSSLKFTKKKNKNYDEAHMPISIGSQG

Ce-UNC-53	1052	TLDNIDVIELKQELKERDSALYEVRLDNLDRAREVDVLRETVNKLKTENKQKKKEVDKLT
yk480b06rc	1025	TLDNIDVIELKQELKERDSALYEVRLDNLDRAREVDVLRETVNKLKTENKQKKKEVDKLT

Ce-UNC-53	1112	NGPATRASSRASIPVIYDDEHVYDAACSSTASQSSKRSSGCNSIKVTNVNDIAGEISSI
yk480b06rc	1085	NGPATRASSRASIPVIYDDEHVYDAACSSTASQSSKRSSGCNSIKVTNVNDIAGEISSI

Ce-UNC-53	1172	VNPDKEIIVGYLAMSTSQCWCWDIDVSILGLFEVYLSRIDVEHQGLIDARDSILGYQIGE
yk480b06rc	1145	VNPDKEIIVGYLAMSTSQCWCWDIDVSILGLFEVYLSRIDVEHQGLIDARDSILGYQIGE

Ce-UNC-53	1232	LRRVIGDSTMITSHPTDILTSSTTIRMFHGAQSRVDSLVLDMLLPKQMILQLVKSIL
yk480b06rc	1205	LRRVIGDSTMITSHPTDILTSSTTIRMFHGAQSRVDSLVLDMLLPKQMILQLVKSIL

Ce-UNC-53	1292	TERRVLGATGIGKSKLAKTLAAYVSIRTNQSEDSIVNISIPENNKEELLQVERRLEKI
yk480b06rc	1265	TERRVLGATGIGKSKLAKTLAAYVSIRTNQSEDSIVNISIPENNKEELLQVERRLEKI

Ce-UNC-53	1352	LRSEKSCIVILDNIPKNRIAFVVSFANVPLQNNEGPFVVCTVNRVYQIPQLQIHHNFKMS
yk480b06rc	1325	LRSEKSCIVILDNIPKNRIAFVVSFANVPLQNNEGPFVVCTVNRVYQIPQLQIHHNFKMS

Ce-UNC-53	1412	VMSNRLEGFILRYLRRRAVEDEYRLTVQMPSELFKIIDFPFIALQAVNNFIEKTNVSDVT
yk480b06rc	1385	VMSNRLEGFILRYLRRRAVEDEYRLTVQMPSELFKIIDFPFIALQAVNNFIEKTNVSDVT

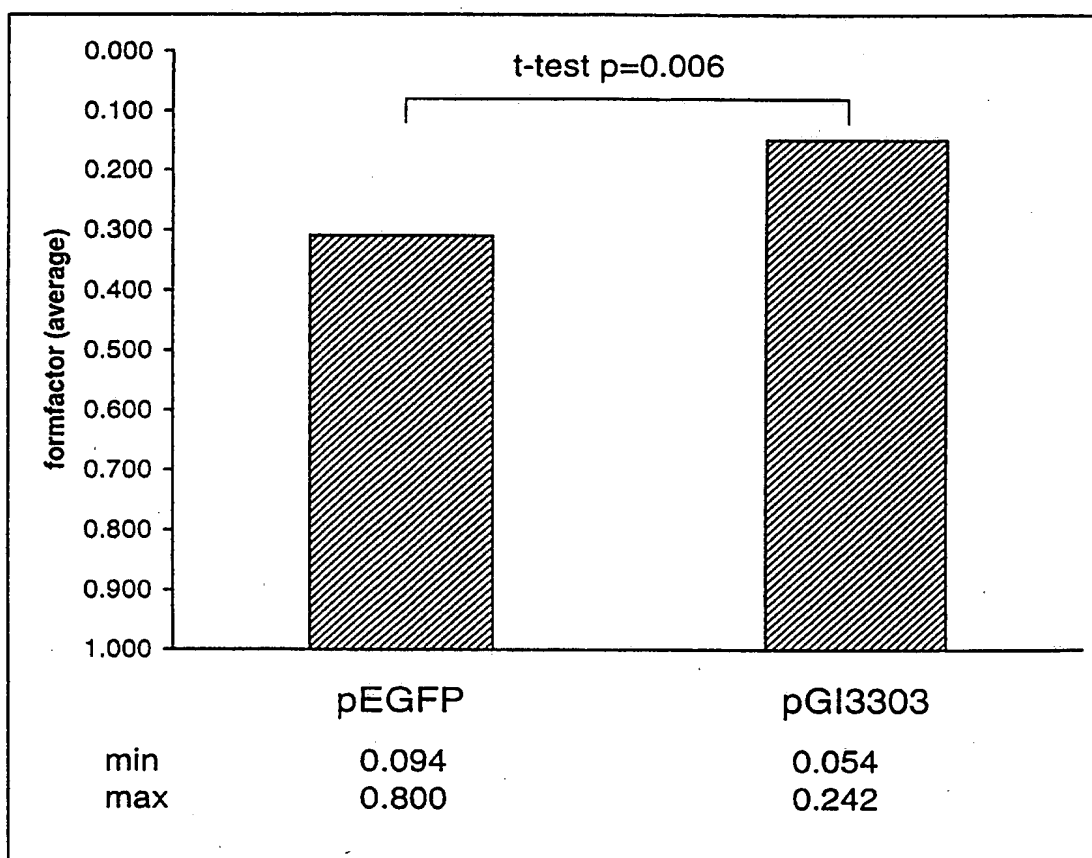
Ce-UNC-53	1472	VGPRACLNCPITVDGSREWFIRLWNENFIPLYLVARVGKKTFGRCTSFEDPTDIVSKKW
yk480b06rc	1445	VGPRACLNCPITVDGSREWFIRLWNENFIPLYLVARVGKKTFGRCTSFEDPTDIVSKKW

Ce-UNC-53	1532	PWFDGENPENVLKRLQLQDLVPSANSSRQHFNPLESILQLHATKHQTIDNI
yk480b06rc	1505	PWFDGENPENVLKRLQLQDLVPSANSSRQHFNPLESILQLHATKHQTIDNI

Legend: the alternative splices and the mutation (S-P) are indicated in red and are boxed.

56/56

Figure 17.



INTERNATIONAL SEARCH REPORT

Interr. appl. No.
PCT/EP 99/03848

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C07K14/47

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 96 38555 A (BOGAERT THIERRY ET AL.) 5 December 1996 (1996-12-05) page 1 -page 99; claims 1-88	1-31, 49-54, 60-67, 72-80, 85,88, 91-95
A	HEKIMI S ET AL: "AXONAL GUIDANCE DEFECTS IN A CAENORHABDITID ELEGANS MUTANT REVEAL CELL-EXTRINSIC DETERMINANTS OF NEURONAL MORPHOLOGY" JOURNAL OF NEUROSCIENCE, vol. 13, no. 10, 1 October 1993 (1993-10-01), pages 4254-4271, XP000612286 ISSN: 0270-6474 the whole document	1,21-26

-/--



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "G" document member of the same patent family

Date of the actual completion of the international search

5 November 1999

Date of mailing of the international search report

22/11/1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

De Kok, A

INTERNATIONAL SEARCH REPORT

International Publication No.

PCT/EP 99/03848

C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	BAIROCH A.: "The PROSITE dictionary of sites and patterns in proteins, its current status" NUCLEIC ACIDS RESEARCH., vol. 21, no. 13, 1993, pages 3097-3103, XP002121559 OXFORD UNIVERSITY PRESS, SURREY., GB ISSN: 0305-1048 the whole document ---	1
P, X	WO 98 24810 A (JANSSEN PHARMACEUTICA) 11 June 1998 (1998-06-11) page 1 -page 97 claims 1-125 ---	1-31, 49-54, 60-67, 72-80, 85,88, 91-95
P, X	NAGASE T ET AL.: "Human mRNA for KIAA0930 protein" EMBL SEQUENCE DATABASE, 9 April 1999 (1999-04-09), XP002121417 HEIDELBERG DE cited in the application Accession Nr.: AB023155 abstract -----	1-11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP 99/ 03848

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☒ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 32-37, 40-45, 49-54, 56, 57, 68-71, 81-84, 86, 87, 89

Present claim 1 relates to an extremely large number of possible vertebrate protein homologues of a UNC-53 protein of *C.elegans*. Support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT is to be found, however, for only a very small proportion of the homologues claimed, i.e. only for the human homologue hs-unc-53/3 (see description page 1, lines 31-34 and page 2, lines 12-15). In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Consequently, the search has been carried out for those parts of the claims which appear to be supported and disclosed, namely those parts relating to hs-unc-53.

Claims 32-37, 40-45, 49-54, 56, 57, 68-71, 81-84, 86, 87 and 89 have not been searched, because they relate to compounds (and their use) whose structural features have not been disclosed at all. Thus, these claims totally lack support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 99/03848

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9638555 A	05-12-1996	AU 6123496 A EP 0832222 A	18-12-1996 01-04-1998
WO 9824810 A	11-06-1998	AU 5662298 A EP 0941239 A	29-06-1998 15-09-1999

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.